

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 10:11:48 ; Search time 2486.45 seconds
(without alignments)
18113.913 Million cell updates/sec

Title: US-09-667-130-1

Perfect score: 3337

Sequence: 1 gaattccggtaaagtaacaa.....tcttataaaataataattc 3337

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_esti:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	435.8	13.1	633	12	AZ569608
2	220	6.6	475	10	BM029212
3	214.2	6.4	691	12	BM040136
4	209	6.3	908	12	AZ548467
5	205	6.1	843	12	AZ551618
6	203.2	6.1	877	12	AZ531291
7	200.6	6.0	890	12	BM146886
8	200.6	6.0	912	12	AZ551092
9	199.8	6.0	931	12	BM160272
10	197	5.9	605	12	AZ640388
11	195.4	5.9	880	12	AZ529191
12	193.8	5.8	735	12	CNS04NSM
13	192.4	5.8	906	12	BM153606
14	192	5.8	501	12	FR0048173
15	191.8	5.7	778	10	BM170064
16	190.4	5.7	905	12	AZ550256
17	189.4	5.7	443	10	BM374465

C 18	186.8	5.6	849	12	AZ546009
19	185.8	5.6	715	10	BM170799
C 20	184.2	5.5	650	12	AZ337339
C 21	182.2	5.5	421	12	AZ113646
22	181.2	5.4	596	12	AZ460541
C 23	180.4	5.4	494	12	FR0048073
C 24	180.4	5.4	816	12	AZ535744
C 25	178.4	5.3	976	12	BM149983
26	178	5.3	891	12	AZ683582
27	176	5.3	942	12	BM148582
C 28	175.8	5.3	879	12	AZ550718
29	175.2	5.3	832	10	BM170146
C 30	173.6	5.2	718	12	AZ972907
31	172.8	5.2	540	12	AZ813205
C 32	171.2	5.1	669	12	AZ392101
C 33	171.2	5.1	898	12	CNS04ALV
C 34	170.4	5.1	773	12	CNS01VTG
C 35	168.4	5.0	824	12	AZ185454
36	167.4	5.0	481	10	BM441017
37	167.2	5.0	677	10	BM161774
C 38	166.8	5.0	619	12	FR0047601
39	166.8	5.0	709	10	BM165622
40	166.4	5.0	617	12	AZ296172
41	166.2	5.0	332	10	BE437454
C 42	166	5.0	774	12	AZ658872
C 43	165.4	5.0	746	10	BM165802
C 44	165	4.9	938	12	CNS07CG2
C 45	164.6	4.9	868	12	BM161951

ALIGNMENTS

RESULT 1

AZ569608 263PVD09 PV MBN #30 Plasmodium vivax genomic 3', DNA sequence.
LOCUS 633 bp DNA linear GSS 15-MAY-2001
DEFINITION 263PVD09 PV MBN #30 Plasmodium vivax genomic 3', DNA sequence.
ACCESSION AZ569608
VERSION AZ569608.1 GI:13979872
KEYWORDS GSS.
SOURCE malaria parasite P. vivax.
ORGANISM Plasmodium vivax
REFERENCE 1 (bases 1 to 633)
AUTHORS Carlton,J.M.-R. and Dame,J.B.
TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL Parasitol. Today 16 (10), 409 (2000)
COMMENT Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: gamej@mail.vetmed.ufl.edu
Seq primer: M13(-20) forward
Class: Shotgun.

FEATURES

Location/Qualifiers
1..633
/organism="Plasmodium vivax"
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69, 497-598)"
/db_xref="taxon:5855"
/clone_lib="pv MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Saimiri boliviensis"
/note="vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V; Host leukocytes were extracted from P. vivax infected blood using the following methods: first, infected blood was activated by the addition of 0.5 ml of ADP (40mg/ml) per 10 ml blood. Then blood was passed over a column of acid washed 0.1 mm glass beads, then through a Plasmidipur filter, followed by passage through a column of pre-wet

Whatman CFl1 powder (1:2 ratio volume of blood to CFl1), and finally centrifuged through a 50% Percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 44% formamide at 50°C as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA Polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pBluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."

```

BASE COUNT      206 a  84 c  196 g  146 t      1 others
ORIGIN

Query Match      13.1%; Score 435.8; DB 12; Length 633;
Best Local Similarity 99.3%; Pred. No. 5.7e-47;
Matches 448; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 261 gatgttagaggtggaggtccttcaaacacgctgttagaattagaagagaaatttt 320
Dy 1 GATGTTGTAGAGGTGGAGGATCCTTCAACACGCGTTTGAATTTAGAAGAGGAAATTTT 60

Qy 321 gatgagaattcagtgatgaaactcttttagatgctaccccccgaagatgactttgcc 380
Dy 61 GATGAGAATTCAGGTGATGATGAACCTCTTTTAGATGCTACCCCGGAGATGACTTTGCC 120

Qy 381 ttaacagatttgccaattgaagcagatgaggaagtcacacgaacgcttagatggagtgaa 440
Dy 121 TTACAGATTTGCCAATTGAAGACGATGAGGAAGTCAACGAAACGTTAGATGGAGGTGAA 180

Qy 441 tcattagagaggtttccactgaagatatgaaacagaagatggtcacaacagatgatacg 500
Dy 181 TCATTAGAGAGGTTTCCACTGAAGATATGGAACAGAAAGATGCTCAACAGATGATACG 240

Qy 501 gaaacagaagaagactacgttgatgataatgaaagagaagaagctggcgatagaa 560
Dy 241 GAAACAGAAGAAGGACTACCTGGTGTATGGAGGAGGAAGAAGCTGGCGATATGGA 300

Qy 561 gcagggaagaagctggtgatttgaagcagggggaacacgctgatttgaagcaggg 620
Dy 301 GCAGGGGAAGAAGCTGTTGATTTGGAGCAGGGGAAGAACTGGCGATTTGGAAGCAGGG 360

Qy 621 gaagaactgctgatttgaagcagggggaagaagctgatttgaagcagggggaaga 680
Dy 361 GAAGAAGCTGCGATTGGAAGCAGGGGAGAGAGCTGCTGATTGGAAGCAGGGGGAAGAA 420

Qy 681 actgcatcttggaagcagggggaagaactg 711
Dy 421 ACTTGGCATTTGG-AGCATGGGAAGAACTG 450

```

```

RESULT 2
BM029212 475 bp mRNA linear EST 05-NOV-2001
LOCUS ipSk00051 Skin cDNA library Ictalurus punctatus cDNA 5', mRNA
DEFINITION
sequence.
ACCESSION BM029212
VERSION BM029212.1 GI:16742782
KEYWORDS EST.
SOURCE channel catfish.
ORGANISM Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
Ictaluridae; Ictalurus.
1 (bases 1 to 475)
Karsi, A., Cao, D., Li, P., Ju, Z., Kocabas, A., Feng, J., Patterson, A.,
Mickett, K.D., and Liu, Z.
Transcriptome analysis of channel catfish (Ictalurus punctatus):
Initial analysis of gene expression and microsatellite-containing
cDNAs in the skin
Unpublished (2001)
JOURNAL
COMMENT Contact: Liu ZJ

```

The Fish Molecular Genetics and Biotechnology Laboratory,
Department of Fisheries and Allied Aquacultures and Program of Cell
and Molecular Biosciences
Auburn University
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA
Tel: 334 844 4054
Fax: 334 844 9208
Email: zliu@acesag.auburn.edu
Seq primer: M13 Reverse.
Location/Qualifiers
1..475
/organism="Ictalurus punctatus"
/db_xref="taxon:7998"
/clone_lib="Skin cDNA library"
/note="Organ: Skin; Vector: pSport1; Site_1: NotI; Site_2:
Sall"

FEATURES
Source

```

BASE COUNT      229 a  77 c  154 g  15 t
ORIGIN

Query Match      6.6%; Score 220; DB 10; Length 475;
Best Local Similarity 69.0%; Pred. No. 3.2e-19;
Matches 301; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 2019 gaattagaagaaccacagagagaagtagaagaacccgtagagggcgagaactgca 2078
Dy 1 GAAGAAGAAGAAGCAGCAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGCAGCAGCA 60

Qy 2079 gaaggagaagaagtgggaaggtaccctgcagaagtagaagaagtggagaagtagctgca 2138
Dy 61 GCAGCAGAAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGAA 120

Qy 2139 gaagtgaagaagtgggaaggtaccagagaagtagaagaaggtaccgcgagaagtagaa 2198
Dy 121 GAACCAGAAGAAGAAGCAGCAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGCAGCA 180

Qy 2199 gaagtgaagaaggtaccagagaagtggaaggtaccagagaagtggaaggtaccac 2258
Dy 181 GCAGCAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGAA 240

Qy 2259 gaagaagtgaagaaggtaccagagaagtggaaggtggagaagtagaagaagtagag 2318
Dy 241 GAAGAAGCAGAAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 300

Qy 2319 gtaccagcggtagtagaagtagaagtagaccagcggtagtagaagaagaggtgccaga 2378
Dy 301 GCAGAAGCAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 360

Qy 2379 gtagaagaagaagaagaaggaaggaaccagtagaagaagaagtagtattacaattagta 2438
Dy 361 GCAGCAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGAA 420

Qy 2439 ataccatcggaagaag 2454
Dy 421 GCAGCAGCAGAAGAAG 436

```

```

RESULT 3
BM040136/c
LOCUS BH040136.1
DEFINITION RPCI-24-255A23.TV RPCI-24 Mus musculus genomic clone RPCI-24-255A23
, DNA sequence.
ACCESSION BH040136
VERSION BH040136.1 GI:14818806
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 691)
Zhao, S., Niernan, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.

```

Qy	1855	aagaggcatttcgaattagaaggaactgcagaagctccagaggaaggaaattagtat	1914
Dd	71	ATGATGATGAAGAAGATGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGAAG	130
Qy	1915	tadaaggaaggaagaccacggaagaagaagccaagcaagggagcgcaacaqaagagag	1974

```
Db 131 AAGATGATGAAGATCAAGATGATGAAGACGATGAAGACGACGAATATCAATTAGAAGATG 190
Qy 1975 aagtcgacagaagaataatgagcgaactccagagagcagatttcgaattagaagaaccaa 2034
Db 191 ATGATGATCAAGACAGAAAGATGATGATGAAGATGATGAAGACGATGAAGACGACGAAT 250
Qy 2035 cagagaagaagttagaagaaccctagagggcggaagaacactgcagaaggagaagaagtgg 2094
Db 251 ATGAATTAGAAGATGATGATGATGAAGAAAGAAAGATGATGATGAAGAAAGAAAGATG 310
Qy 2095 aagagggtacctgcagaagttagaagaagtgggaagggtacctgcagaagttagaagaagtgg 2154
Db 311 ATGATGAAGATGATGATGAAGAAAGAAAGATGATGAAGAAAGAAAGATGATGAAGAAAG 370
Qy 2155 aagagggtaccagaagaagttagaagaagtcaccgcagaagtagaagaagtaggaaggtac 2214
Db 371 ATGATGAAGAAAGAAAGATGATGAAGATGATGAAGACGATGAAGACGACGAATATG 430
Qy 2215 cagaagaagtgggaaggtaccagaagaagtgggaaggtaccagaagaagtgggaaggtgg 2274
Db 431 AATTAGAAGATGATGATGAAGAAAGAAAGATGATGAAGAAAGAAAGATGATGATGATG 490
Qy 2275 taccagaagaagtgggaagaagttagaagaagtagaagaagtagaggtaccagcggtagtag 2334
Db 491 AAGATGATGAAGAAAGAAAGATGATGAAGATGATGAAGAAAGAAAGATGATGAAGATGATG 550
Qy 2335 aagtagaagtagcaccagcgttagaagaagaaggtgcccagaagaagtagaagaagaagaag 2394
Db 551 AAGAAAGAGATGATGATGAAGAAAGAAAGATGATGAAGAAAGAAAGATGATGAAGAAAGATG 610
Qy 2395 aagagggaagaccagtaggagaagaagttagtaccattagtagtaataaccatcggaagaag 2454
Db 611 AAGAAAGAAAGATGATGATGAAGACGACGAATATGAATTAGAAGATGATGAAGAAAGATG 670
Qy 2455 atatacattagacaacccaagaagaagcgaattaggctctggaattttatctatcatcg 2514
Db 671 ATGAAGAGATGATGATGAAGAAAGATGATGAAGAAAGAAAGATGATGAAGATGATGATG 730
Qy 2515 acatgcactaccagacgttccaaaggaatttatggagaagaagaagaagaactcagtggt 2574
Db 731 AAGAAAGATGATGATGAAGATGATGAAGAAAGATGATGAAGATGATGAAGATGATGAAGATG 790
Qy 2575 atccattgaaaccagaagattttccaaagaagaagattcacattctacaga 2623
Db 791 AAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATG 839

RESULT 5
LOCUS A2551618 843 bp DNA linear GSS 14-NOV-2000
DEFINITION ENT54TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION A2551618
VERSION A2551618.1 GI:11176919
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 843)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@loftusetigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
```

```
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 39
High quality sequence stop: 838.
FEATURES
    source
        1..843
            /organism="Entamoeba histolytica"
            /strain="HMI:IMSS"
            /db_xref="taxon:5759"
            /clone_lib="Entamoeba histolytica Sheared DNA"
            /note="Vector: pHOS1; Site:1; Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999)."
BASE COUNT 155 a 266 c 18 g 404 t
ORIGIN
Query Match 6.1%; Score 205; DB 12; Length 843;
Best Local Similarity 55.6%; Pred. No. 2.3e-17;
Matches 394; Conservative 0; Mismatches 315; Indels 0; Gaps 0;
Qy 1855 aagaggcagatttcgaattagaagaactgcagaagtcacagaggaaggagaattagtagt 1914
Db 807 ATGATGATGAAGAAAGATGATGATGAAGAAAGATGATGAAGAAAGATGATGAAGAAAGATG 748
Qy 1915 tagaagagaagaagaacccaacggaagaagcccaagaagaagaagagcccaacaagaagag 1974
Db 747 AGTATGATGAAGATGAAGATGATGAAGACGATGAAGACGATGAAGATGAAGATGAAGATG 688
Qy 1975 aagtcgacagaagaataatgagcgaactccagagagcagatttcgaattagaagaaccaa 2034
Db 687 ATGATGATGAAGAAAGAAAGATGATGATGAAGATGATGAAGACGATGAAGACGACGAAT 628
Qy 2035 cagagaagaagtagaagaacccctagagggcggaagaacactgcagaaggagaagaagtgg 2094
Db 627 ATGAATTAGAAGATGATGATGATGAAGAAAGAAAGATGATGAAGAAAGAAAGATGATG 568
Qy 2095 aagagggtacctgcagaagaagaagtgggaaggtacctgcagaagtagaagaagtgg 2154
Db 567 ATGATGAAGATGATGATGAAGAAAGAAAGATGAAGAAAGATGATGAAGAAAGATGATGAAG 508
Qy 2155 aagagggtaccagaagaagtagaaggttaccgcagaagtagaagaagtaggaaggttag 2214
Db 507 ATGATGAAGAAAGAAAGATGATGATGAAGATGATGAAGACGATGAAGACGACGAATATG 448
Qy 2215 cagaagaagtgggaagaggtaccagaagaagtgggaaggtaccagaagaagtaggaagaag 2274
Db 447 AATTAGAAGATGATGATGAAGAAAGAAAGATGATGAAGAAAGAAAGATGATGAAGATGATG 388
Qy 2275 taccagaagaagtgggaagaagtagaagaagtagaagaagtaggaaggttagcagcggttag 2334
Db 387 AAGATGATGAAGACAGAAAGATGAAGATGATGATGAAGAAAGAAAGATGAAGATGATGATG 328
Qy 2335 aagtagaagtagcaccagcgttagaagaagaaggtgcccagaagaagtaggaagaagaagag 2394
Db 327 AAGAAGAAAGATGATGATGAAGAAAGAAAGATGAAGAAAGAAAGATGATGAAGAAAGATGATG 268
Qy 2395 aagagggaagaaccagtagaggaagaagtagtattacaattagtaataaccatcggaagaag 2454
Db 267 AAGAAGAAAGAAAGATGATGATGAAGACGACGAATATGAATTAGAAGATGATGAAGAAAGATG 208
Qy 2455 atatacattagacaacccaagaagaagcgaattaggctctggaattttatctatcatcg 2514
```



```
Db 207 ATGAAGAAGATGATGAAGAAGAGATGATGAAGAAGAGATGATGATGATGATG 148
QY 2515 acatgcactaccagcgttcccaaggaatttatggaagaagaagaaga 2563
Db 147 AACAAAGATGATGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAG 99

RESULT 6
AZ531291/c 877 bp DNA linear GSS 03-NOV-2000
LOCUS ENTBQ347R Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic DNA sequence.
ACCESSION AZ531291
VERSION AZ531291.1 GI:11085838
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 877)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 22
High quality sequence stop: 829.
Location/Qualifiers
1..877
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1; Site1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

BASE COUNT
ORIGIN

```
Query Match 6.1%; Score 203.2; DB 12; Length 877;
Best Local Similarity 54.9%; Pred. No. 3.8e-17;
Matches 400; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

QY 1749 gatgcaggagggctgtacagctaccataaaggagccacctgtacaagtaccaggga 1808
Db 815 GAAGATGAAGACCAATATGATTAAGAGAGAGAGAGATGATGATGAAGAAGAA 756
QY 1809 gtaggccgcgcgaagaagtgcgaacgggaagaattgatgcaactccaaggagcgatttc 1868
Db 755 GAAGATGATGAAGATGATGATGAAGAAGATGATGATGAAGAAGATGATGATGAAGAA 696
QY 1869 gaattagaaggaaactcagaagctccaggagggaagggaatttagtattagaaggaga 1928
```

FEATURES

Location/Qualifiers

```
Db 695 GATGATGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGAGATGATGAAGAT 636
QY 1929 gaaccaacggaagaagaagccaaagagagagagccaaacagaagaagaagtgccagaagaa 1988
Db 635 GAAGATGATGAAGACCGATGAAGACGACGAATATGATTAAGATGATGATGAAGAA 576
QY 1989 gaattagaggcaactccagaggacgatttcgaatttagaagaacacacaggagaagta 2048
Db 575 GAAGAAGATGATGAAGATGATGAAGACGATGAAGACGATGAAGACGACGAATATGAATTAGAAGAT 516
QY 2049 gaagaacacctagaggcgcaagaagaactgcgaaggagaagaagtgggaagaggtacctgcga 2108
Db 515 GATGATGATGATGAAGAAGAGATGATGATGAAGAAGAGATGATGATGAAGATGAT 456
QY 2109 gaagtgaagaagtgggaagggttacctgcagaagtagaagaagtgggaagggtaccagaa 2168
Db 455 GATGAAGAAGACATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGAA 396
QY 2169 gaagtgaagaagggtaccgcgcagaaagtagaagaagtggaaggtaccagagaagtgga 2228
Db 395 GAAGATGATGATGAAGATGATGAAGACGATGAAGACGACGAATATGAATTAGAAGATGAT 336
QY 2229 gaggtaccagaagaagtgggaagggtaccagaagaagtgaagaaggtaccagaagaagt 2288
Db 335 GATGAAGAAGAGATGATGATGAAGAAGAGATGATGAAGAAGATGATGAAGATGATGAAG 276
QY 2289 gaagaagtgggaagaagtagaaggtagaggtaccagcggtagtagaagtagaagtagacca 2348
Db 275 GAAGAAGATGAAGATGATGATGAAGAAGAGATGAAGATGAAGATGAAGATGAAGATGAT 216
QY 2349 gcggttagtagaagaagaaggtgcccagaagaagtagaagaagaagaagaaggaagaaacca 2408
Db 215 GATGAAGAAGAGATGAAGAAGAGAGATGATGAAGAAGATGATGAAGAAGAGAGAT 156
QY 2409 gttagagaagaagattgattacattagtaataccatccgcgaagaagatatcaattagac 2468
Db 155 GATGATGAAGACCGACGAATATGAATTAGAAGATGATGAAGAAGATGATGAAGAAGATGAT 96
QY 2469 aaacacaaa 2476
Db 95 GAAGAAGA 88

RESULT 7
BHI46886
LOCUS ENTBQ48TF Entamoeba histolytica Sheared DNA linear GSS 27-AUG-2001
DEFINITION genomic DNA sequence.
ACCESSION BHI46886
VERSION BHI46886.1 GI:15302963
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 890)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 880.
Location/Qualifiers
```

```

source
1. .890
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: PHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      404 a   65 c   285 g   136 t
ORIGIN

Query Match      6.0%;   Score 200.6;   DB 12;   Length 890;
Best Local Similarity 54.7%;   Pred. No. 8.2e-17;
Matches 398;   Conservative 0;   Mismatches 329;   Indels 0;   Gaps 0;

QY 1741 aggtctctgacgagagcgtgaacagtcacatcaaggagcagcctgtacaagtac 1800
      ||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 ACGATGATGATGAAGATGACGATGATGATGAAGATGATGATGACGATGAAGACG 137

QY 1801 cagtggcagtagggccgcgaagaagtgcacacggagaattgattgcactccaagg 1860
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 ATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACG 197

QY 1861 acgatttcgaattagaaggaactcagaagctccagaggaaggaattagattacaag 1920
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 ATGAAGAAGACGATGAAGAAGACGATGATGAAGAAGACGATGATGAATTTGAATTAGAAG 257

QY 1921 gagaaggagaacacacggaagaagcgaagagagagagcgaacagagaaggaggaagtgc 1980
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 ACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAG 317

QY 1981 cagaagaagaattagaggaactccagagagcagattcttcgaattagaagaacacagagg 2040
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 ACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAG 377

QY 2041 aagaagtagaagaacacgtagagggccgaacaaactcagaagagagaagaagtggaaagg 2100
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 ACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAG 437

QY 2101 tacctgcagaagttagaagaagtggagaggtaccctgcagaagtagaagaagtggaaagg 2160
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 ACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAG 497

QY 2161 taccagaagaagttagaagaaggtaccgcagaagtagaagaagtggaggtaccagaag 2220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 498 ATGAAGAAGATGAAGATGATGATGATGAAGAAGACGATGATGAAGAAGATGATGAAGAAG 557

QY 2221 aagtgggaagggtaccagaagaagtgaagaaggtaccagaagaagtggaaagggtaccag 2280
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 ATGAAGACGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACG 617

QY 2281 aagaagtgggaagaagttagaagaagttagaggtaccagcggtagtagaagaagtag 2340
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 ATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACG 677

QY 2341 aagtaccagcggtagtagaagaagaggtgcccagaagaagttagaagaagaagaagagg 2400
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 ATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACG 737

QY 2401 aagaacacagtagaagaagattattacaattagtaattaccatcggagaagaagatatc 2460
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 ATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAATTTG 797

```

```
Qy 1901 aggaagaattagtagtattagaagaagagagagaaacccggaagaagagccaagagagagaga 1960
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 AGAAGATGATGAAGCAAGCAAGAGATGATGAAGATGAAGATGATGAAGATGAAGATGAAGATGA 388

Qy 1961 gccaacgaaggaaggaagtcgcaagaagaattagaggaactccagagagcagattccga 2020
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 389 CGAATATGAATTAGAAATGATGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 448

Qy 2021 attagaagaaccaacaggaagaagaagtagaagaacccgtagagggcgcaagaacctgcaga 2080
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 449 AGACGATGAACACACCAATATGAATTAGAAGATGATGATGAAGATGAAGATGAAGATGAAGATGA 508

Qy 2081 aggaagaagaagtggaaggtaccctgcagaagtagaagaagtggaaggtaccctgcaga 2140
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 509 TGATGAAGAGAGCAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGA 568

Qy 2141 agtagaagaagtggaaggtaccagagaagtagaagaaggtaccctgcagaagtagaaga 2200
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 569 TGAAGAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGA 628

Qy 2201 agtggaagaaggtaccagagaagtggaaggtaccagagaaggtaccagagaaggtaccagaga 2260
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 CGATGAAGAGACCAATATGAATTAGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGA 688

Qy 2261 aggaagtggaaggtaccagagaagtggaaggtaccagagaaggtaccagagaaggtaccagaga 2320
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 689 AGAAGAGAGAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGA 748

Qy 2321 accagcggtagtagaagtagaaggtaccagcggttagtagaagaaggtgacagagaaggt 2380
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 749 AGAAGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGA 808

Qy 2381 aggaagaagaagaagaagaaccagtagaggaagaagaagtagtagaagaagtagtagaagaat 2440
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 809 AGATGATGAAGAGATGATGAAGAGAGAGAGATGATGATGAAGAGAGAGATGATGATGAAGAGAGAGATGA 868

Qy 2441 accatcggaagaaga 2455
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 869 TTAAGATGATGAAGA 883

RESULT 9
LOCUS BH160272
DEFINITION ENTQV49TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION BH160272
VERSION BH160272.1 GI:15733710
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 931)
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library (2001)
Unpublished (2001)
JOURNAL
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 8
High quality sequence stop: 677.
FEATURES
source location/Qualifiers
1..931
```

/organism="Entamoeba histolytica"

/strain="HM1:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 402 a 51 c 279 g 199 t
ORIGIN

Query Match 6.0%; Score 199.8; DB 12; Length 931;
Best Local Similarity 55.6%; Pred. No. 1e-16;
Matches 384; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

Qy 1763 tgtaacagtagtaccatcaagaagcaccctgtacaagtagcagtagggccgcgca 1822
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1 TCGAAGATTGACAGATGATGATGAAGAGAGAGAGATGATGATGAAGATGATGAAGA 60

Qy 1823 agaagtgccaacggaagaattgatgcaactccaagagagcatttcgaattagaaggaac 1882
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 61 CGATGAAGACGACCAATATGAATTAGAAGATGATGATGAAGAGAGAGATGATGA 120

Qy 1883 tgcaagaagtcctccaggaaggaaggaattagtagtagagaggaaggaaggaaggaagga 1942
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 121 TGAAGAGAGAGAGAGATGATGATGAAGATGATGATGAAGAGAGAGATGATGAAGAAGATGATGA 180

Qy 1943 agagccaagagaagagagcccaacagaggaagtgccagagaagaataatagaggcaac 2002
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 181 AGAAGATGATGAAGATGATGATGAAGAGAGAGAGATGATGATGAAGATGATGAAGACGA 240

Qy 2003 tccagaggaaggttcgaattagaagaaccacaggaagaagtagaagaacacgtaga 2062
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 241 TGAAGACGACCAATATGAATTAGAAGATGATGATGAAGAGAGAGATGATGAAGAAGATGATGAAGA 300

Qy 2063 gggcgagaagactgcagaaggaagaagtggaagaggtaccctcagaagtagaagaagt 2122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 301 AGAAGAAGATGATGATGAAGATGATGAAGAAGAGATGATGAAGATGATGAAGATGAAGAAGA 360

Qy 2123 ggaagaggtacctgcagaagttagaagaagtgggaagaggtaccagagaagaagtagaagaggt 2182
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 361 AGATGAAGATGATGATGAAGAAGAGATGATGATGAAGAAGAGATGATGAAGAAGAGAGAGA 420

Qy 2183 acccgagaagtagaagaagtgggaagaggtaccagagaagaagtggaagaggtaccagaga 2242
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 421 TGATGAAGAAGATGATGAAGAAGAGAGATGATGATGAAGACGACCAATATGAATTAGA 480

Qy 2243 agtggaaggtaccagagaagaagtgggaaggtaccagagaagaagtggaagaggtggaaga 2302
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 481 AGATGATGAAGAAGATGATGAAGAAGAGATGATGAAGAAGAGATGATGAAGAAGAGAGAGA 540

Qy 2303 agtagaagaagtagaggtaccagcggttagtagaagaagtagaccagcggttagtagaaga 2362
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 541 TGATGATGATGATGATGAAGAAGATGATGATGAAGAAGATGATGAAGAAGATGAAGATGAAGA 600

Qy 2363 agaggtgccgaagaagtagaagaagaagaagaagaagaaggaaggaaggaaggaaggaaga 2422
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 601 AGAAGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGAAGA 660

Qy 2423 tgattacaattagtaataaccatcggaaga 2453
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 661 CGCATTATGATTAGAAGAGATGAGATGAAGA 691

```
RESULT 10
AZ640388/c 605 bp DNA linear GSS 14-DEC-2000
LOCUS
DEFINITION
clone UUGCLM0502E21 F, DNA sequence.
ACCESSION
AZ640388
VERSION
AZ640388.1 GI:11763788
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (Bases 1 to 605)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid Inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0502 row: E column: 21
Seq primer: CGTGTAAACGACGCGGCAGT
Class: plasmid ends
High quality sequence stop: 605.
Location/Qualifiers
1..605
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0502E21"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g114732114[gblAF129072.1]), a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
13 a 253 c 136 g 203 t

BASE COUNT
ORIGIN

Query Match 5.98; Score 197; DB 12; Length 605;
Best Local Similarity 59.38; Pred. No. 2.7e-16;
Matches 335; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 1874 agaaggaactcagaagtcagaggaaggaatttagtattagaagagagaacc 1933
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 570 AGCAGGACGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 511
```

```
Qy 1934 aacggaagaagagccaagagagagagccaacagaagaagtgccagaagaagt 1993
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 510 AGCAGGACGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGG 451
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1994 agaggaactcagagagcagatttcgaattagaagaacacagagagaagtagaaga 2053
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 450 AGCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGG 391
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2054 aaccgtagaggcggaagaactgcagaagagagaagaagtgcacctgcagaagt 2113
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 AGCAGCAGCAGCAGCAGCAGGCGGCGAGCAGGCGGCAGCAGCAAGAGCAGCAGCAGG 331
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2114 agaagaagtggaaagggttacctgcagaagtagaagaagtggaaagggttaccagaagaagt 2173
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 330 AGCAGCAGGCGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGGCAGCAGCAGCAGA 271
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2174 agaagaggttaaccgcagagaagttagaagaagtggaaagggttaccagaagaagtggaaagagt 2233
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 270 AGCAGCAGGCGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGAGCAGCAGGAGCAGTAGGAGCAGTAGG 211
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2234 accagaagaagtggaaagggttaccagaagaagtggaaagggttaccagaagaagtggaaaga 2293
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 AGCAGTAGGAGCAGTAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGCAGC 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2294 agtggaaagaagttagaagaagttagagggttaccagcggttagtagaagttagaagtagaccagcggt 2353
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 AGGAGCAGCAGAGAAGCAGCAGGCGGCGAGCAGCAGCAGCAGCAGCAGCAGGCGGCAGCAGCAGC 91
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2354 agtgaagaagaagtggccagaagaagttagaagaagaagaagaagaagaagaccagtagta 2413
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 90 AGCAGACGACGAGCAGTAGGAGCAGTAGGAGCAGTAGGAGCAGCAGCAGCAGCAGCAGGAGGAGC 31
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2414 ggaagaagatgtattacaattagta 2438
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30 AGCAGGACGACTAGGAGCAGCAGTAGGA 6
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AZ529191/c 880 bp DNA linear GSS 03-NOV-2000
LOCUS
DEFINITION
ENTBV68TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION
AZ529191
VERSION
AZ529191.1 GI:11081835
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica.
ORGANISM
Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
1 (Bases 1 to 880)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL
Unpublished (2000)
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 673.
Location/Qualifiers
1..880
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"

FEATURES
source
```


[illegible]

```

RESULT 13
BH153606/c
LOCUS
DEFINITION
    BH153606 linear GSS 24-SEP-2001
    ENT83TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
    genomic, DNA sequence.
ACCESSION
    BH153606
VERSION
    BH153606.1 GI:15725323
KEYWORDS
    GSS.
SOURCE
    Entamoeba histolytica.
    Entamoeba histolytica
    Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
    1. (bases 1 to 906)
    Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
    Determination of clone end sequences from Entamoeba histolytica
    HMI:IMSS sheared DNA library (2001)
    Unpublished (2001)
JOURNAL
    Contact: Brendan J Loftus
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850, USA
    Tel: 301 838 0208
    Fax: 301 838 3543
    Email: b.loftus@tigr.org
COMMENT
    Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
    DNA library
    Seq primer: M13-Forward
    Class: shotgun
    High quality sequence start: 15
    High quality sequence stop: 733.
    Location/Qualifiers
        1..906
FEATURES
    source

```

```

i. 3906
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica sheared DNA"
/notes="Vector: pHOs1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999).".
149 a 293 c 66 g 398 t

```

	Query Match	5.98;	Score 192.4;	DB 12;	Length 906;
	Best Local Similarity	55.18;	Pred. No. 9.2e-16;		
	Matches 376;	Conservative	0;	Mismatches 306;	Indels 0; Gaps 0;
QY	1743	gatcctgatgcagagaggcgtgaacagtcacatcaaggaagcacctgtacaagtacca	1802		
DB	733	gatgatgaagatgacgatgatgatgcagaagatgatgatgccgatgacgaacacgat	674		

Qy	1803	gtggcagtagggccgcgcgaagtcgccaacggaagaaattgtgcaactccaagagac	1862
Db	673	GAAGAACACGATGATGAAGACCGATGATGAAGACCGATGAAGAACGATGAAGAACGAT	614
Qy	1863	gatttcaattagaagaaactcagaagctccagagggaagagaattagattagaagga	1922
Db	613	GAAGAGCAGCGATGATGAAGACGATGATGAAGAACGACGATGATGAATTTGAATTAAGAC	554
Qy	1923	gaaggagaaccaacgcgaagaagagccaagagagagagccaacagaagaggaagtgcga	1982
Db	553	GATGATGAAGAACGACGATGATGAAGAACGACGATGATGAAGAACGACGATGATGAAGAAC	494
Qy	1983	gaagaagaattagaggccaactccagaggagcagatttcgaattagaagaacccaacggagaa	2042
Db	493	GATGATGAAGAACGACGATGATGAAGAACGACGATGATGAAGAACGACGATGATGAAGAAC	434
Qy	2043	gaagtagaagaaccgttagagggcgaagaaacctgcagaaggagagaagtgtgaagagta	2102
Db	433	GATGATGAAGAACGACGATGATGAAGAACGACGATGATGAAGAACGACGATGATGAAGAAC	374
Qy	2103	ccctgcagaagttagaagaagtggaaaggttacctgcagaagtagaagaagtggaaagagta	2162
Db	373	GATGATGAAGAACGACGATGATGAAGAACGACGATGATGAAGAACGACGATGATGAAGAAC	314
Qy	2163	ccagaagaagttagaagaggttaccgcagaagtagaagaagtggaaaggttaccagaagaa	2222
Db	313	GAGATGATGAAGATGATGATGATGAAGAACGACGATGAAGAACGATGATGAAGAACGACG	254
Qy	2223	gtggaagaggttaccagaagaagtggaaaggttaccagaagaagtggaaaggttaccagaa	2282
Db	253	GAAGACGACGATGAACACGACGATGAAGACGACGATGAAGAACGACGATGAAGAACGAT	194
Qy	2283	gaagtggaagaagtggaaagaagttagaagaagttagagggttaccagcggtagtagaagtaa	2342
Db	193	GAAGAAGACCGATGAAGAAGACCGATGAAGAAGACCGATGAAGAAGACCGATGAAGAAGAC	134
Qy	2343	gtaccagcggtagtagaagaagaagtgcagaagaagttagaagaagaagaagaagagtaa	2402
Db	133	GAGAACGACGATGAAGAAGACCGATGAAGAACGACGATGAAGAACGACGATGAAGAACGAT	74
Qy	2403	gaaccagtagaaggaagaagatg	2424
Db	73	GAAGAAGACCGATGAAGAAGACG	52

RESULT	14				
LOCUS	FR0048173/c				
DEFINITION	Fugu rubripes GSS sequence, clone 264E22ca9, genomic survey				
ACCESSION	AL44958				
VERSION	AL44958.1	GI:12053458			
KEYWORDS	GSS; genome survey sequence.				
SOURCE	Takifugu rubripes.				
ORGANISM	Takifugu rubripes				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Actinopterygii; Neopterygii; Telostei; Euteleostei;				
	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;				
	Tetraodontidae; Takifugu.				
REFERENCE	1	(bases 1 to 501)			

REFERENCE	1 (bases 1 to 501)
AUTHORS	Clark, M.S.
TITLE	Direct Submission
JOURNAL	Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. UK Email: biohelp@hgmrc.ac.uk
COMMENT	Vector: pBluescript II KS V_type: phagemid PRIMER: KS DESCR:

DESCRIPTION:
One pass dye-terminator sequencing of BAC (pBelobACII) cloned genomic sequence
The BACs can be obtained from <http://www.incyte.com>.
Location/Qualifiers

FEATURES

[illegible]

JOURNAL
COMMENT

Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0068 row: B column: 03
Seq primer: CGTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 650.
Location/Qualifiers
1. .650
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0068B03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: FWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (g114732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

FEATURES

source

1. .650
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0068B03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: FWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (g114732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT
ORIGIN

Query Match 5.5%; Score 184.2; DB 12; Length 650;
Best Local Similarity 57.3%; Pred. No. 1.2e-14;
Matches 333; Conservative 0; Mismatches 248; Indels 0; Gaps 0;
Qy 1874 agaaggaactgcagaagctccagaggaggaattagattaggaaggaaggaacc 1933
Db 649 AGCAGGAGGAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 590
Qy 1934 aacggaagagagcagaagagagagccacagaaggaggaagtgcacagaagaagaatt 1993
Db 589 AGGAGCAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 530
Qy 1994 agaggcaactccagagacatttcgaattagaagaacacaggaagaagaattagaaga 2053
Db 529 AGCAGGAGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 470
Qy 2054 aaccgtgagggcgcaagaactgcagaaggaagaagtggaagaggtacctgcagaagt 2113
Db 469 AGGAGGAGCAGGAGCAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 410
Qy 2114 agaagaagtggaagaggtacctgcagaagtagaagaagtggaagaggtaccagaagaatt 2173
Db 409 AGGAGGAGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 350
Qy 2174 agaagaggtaccgcagaagtagaagaagtgaagaaggtaccagaagaagtggaagaggt 2233
Db 349 AGCAGGAGCAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 290

QY 2234 accagaagaagtgaagaggtaccagaagaagtgaagaggtaccagaagaagtgaaga 2293
Db 289 AGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 230
QY 2294 agtgaagaagtagaagaagttagaggtaccagcggttagttagaagttagaagtagaccagcggt 2353
Db 229 AGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 170
QY 2354 agtagaagaagaagtgccagaagaagttagaagaagaagaagaagaagaaccagtaga 2413
Db 169 AGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 110
QY 2414 ggaagaagatgtattacaattagtaataaccatcggaagaag 2454
Db 109 AGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 69

RESULT 21
AZ113646/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
AZ113646
421 bp DNA linear GSS 12-MAY-2000
RPCI-23-447D2.TJ RPCI-23 Mus musculus genomic clone RPCI-23-447D2
DNA sequence.
AZ113646
GI:7773245
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 421)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Aklnret
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-447D2.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 447 row: D column: 2
Seq primer: SP6
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. .421
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-447D2"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1;
ECORI; Site:2: EcorI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcorI and EcorI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
15 a 134 c 24 g 248 t

BASE COUNT
ORIGIN

```
Query Match      5.5%; Score 182.2; DB 12; Length 421;
Best Local Similarity 68.9%; Pred. No. 2.4e-14;
Matches 250; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 2039 agaagaagttagaagaaccgtagagggcggaagaactgcagaaggaagaagtgggaaga 2098
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 417 AGACGACGGGGAAGACGAGTTCGACGCGGAGGAAGGAAGGAAGGAAGGAAGGAAGA 358

Qy 2099 ggtacctgcagaagttagaagaagtgggaaggttacctgcagaagttagaagaagtgggaaga 2158
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 AGAAGCAGAACGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 298

Qy 2159 ggtaccagaagaagttagaagaaggttaccgcagaagtggaagaagtgggaaggttaccaga 2218
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 297 AGAAGAAGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 238

Qy 2219 agaagtggaaggttaccagaagaagtgggaaggttaccagaagaagtgggaaggttacc 2278
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 237 ACAAGAGNAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 178

Qy 2279 agaagaagtgggaagaagttagaagaagtagaggttaccagcggttagtagaagt 2338
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 177 AGAAGAAAAAAGCAATTTGTAGTGAAGAAGAAGAGTAAAGAAGAAGAAGAAGA 118

Qy 2339 agaagttaccagcggttagtagaagaagaaggttccagaagaagttagaagaagaagaaga 2398
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 117 AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 58

Qy 2399 gga 2401
      ||
Db 57 GAA 55

RESULT 22
AZ460541
LOCUS      AZ460541
DEFINITION 1M0265I24R Mouse 10kb plasmid UUGCIM library Mus musculus genomic
clone UUGCIM0265I24 R, DNA sequence.
ACCESSION  AZ460541
VERSION     AZ460541.1 GI:10618666
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 596)
            Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
            Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: dduenne@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0265 row: I column: 24
            Seq primer: CACACAGGAACACGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 596.
            Location/Qualifiers
                1..596
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UUGCIM0265I24"
                    /clone_lhb="Mouse 10kb plasmid UUGCIM library"

FEATURES
source
```

```
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gil4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      315 a      59 c      145 g      77 t
ORIGIN

Query Match      5.4%; Score 181.2; DB 12; Length 596;
Best Local Similarity 70.2%; Pred. No. 2.9e-14;
Matches 257; Conservative 0; Mismatches 108; Indels 1; Gaps 1;

Qy 2036 agagaagaagttagaagaaccgttag-agggcggaagaactgcagaaggaagaagtgg 2094
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 AGAAGAACAAGAAGAACAAGGCAAGCAAGAAGAAGAACAAGAAGAACAAGAACAAG 61

Qy 2095 aagagggtacctgcagaagttagaagaaggttacctgcagaagttagaagaagtgg 2154
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 AAGAGAACAAGAAGAAGAAGAACAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 121

Qy 2155 aagagggtaccagaagaagttagaaggttaccgcagaagtagaagaagtggaaggttac 2214
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 AAGAAGAAGAAGAAGAAGAACAACAAAAGAACAAGAACAAGAACAAGAAGAAG 181

Qy 2215 cagaagaagtgggaaggttaccagaagaagtggaaggttaccagaagaagtgggaaggg 2274
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 AAGAAGAACAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 241

Qy 2275 taccagaagaagtgggaaggttagaagaagtagaagaaggttaccgcaggttagtag 2334
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 AAGACAAGAAGAAGAACAAGAAGAAGAAGAAGAACAAGAAGAACAAGAAGAAGAAG 301

Qy 2335 aagtagaagtaccagcggttagtagaagaaggtgcccagaagaagtagagaagaagaag 2394
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 361

Qy 2395 aagagg 2400
      |||||
Db 362 AAGAGG 367

RESULT 23
FR0048073/c
LOCUS      FR0048073/c
DEFINITION Fugu rubripes GSS sequence, clone 263K15bD8, genomic survey
sequence.
ACCESSION  AL444858
VERSION     AL444858.1 GI:12052694
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes
ORGANISM    Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
            1 (bases 1 to 494)
```

[illegible]

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjlouftus@tigr.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared

DNA library

Seq primer: M13-Reverse

Class: shotgun

High quality sequence start: 16

High quality sequence stop: 694.

Location/Qualifiers

FEATURES

source

/organism="Entamoeba histolytica"

/strain="HM1:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOS1; Site_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999)."

BASE COUNT 400 a 55 c 291 g 145 t

ORIGIN

Query Match 5.3%; Score 178; DB 12; Length 891;

Best Local Similarity 54.2%; Pred. No. 6.6e-14;

Matches 361; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

QY 1790 tgtacaagttaccagtgccagtagggcccgcaagagagtgccacggaagaattgtatgca 1849

DB 58 TGAAGATGATGAATTTGCAATTAGAACGCGATGATGAAGATGATGAAGATGATGAAGATGATGA 117

QY 1850 actccaagagcagatttcgaattagaagaactcagaagctccagagagcagaagaatt 1909

DB 118 TGAAGATGACGATGATGAAGAACGATGATGAAGAACGATGATGAAGAACGATGATGAAGAACGATGA 177

QY 1910 agtattagaaggaaggaagaaacacggaagagagccaggaagagagagagcacaaga 1969

DB 178 TGATGAAGAGATGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 237

QY 1970 aggagaagtgccagaagaagaattagaggcaactccagaggagcagatttcgaattagaaga 2029

DB 238 AGATGATGATGATGATGAAGAACGATGAAGAACGATGAAGAACGATGAAGAACGATGAAGAACGATGA 297

QY 2030 accaagaggaaggaagtaagaagaacccgtagagggcggaagaacacgcaagaggaaga 2089

DB 298 CGATGAAGAACGATGAAGAACGATGAAGAACGATGAAGAACGATGAAGAACGATGAAGAACGATGA 357

QY 2090 agtgggaaggtaccctgcagaagtagaagaagtgggaagaggtaccctgcagaagtagaaga 2149

DB 358 AGACGATGATGATTTGCAATTAGAACGCGATGATGAAGAACGCGATGAAGAACGCGATGAAGAACGCGA 417

QY 2150 agtgggaaggtaccagaagaagtagaagaaggtaccgcgcaagtagaagaagtggaaga 2209

DB 418 TGATGAAGAACGATGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGAAGAACGCGA 477

QY 2210 ggtaccagaagaagtggaagaggtaccagaagaagtgggaagaggtaccagaagaagtggga 2269

DB 478 TGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGAAGAACGCGA 537

QY 2270 aggggtaccagaagaagtggaagaagtagaagaaggtaccagaaggtaccagcgggt 2329

DB 538 TGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGAAGAACGCGA 597

QY 2330 agtagaagtagaagttaccagcggtagtagaagaagaggggtccagaagaagtagaagaaga 2389

DB 598 TGATGAAGAACGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGA 657

QY 2390 agaagaaggggaagaccagtagtagaagaagaagatgtattacaattgtaataccatcgga 2449

DB 658 TGAAGACGACGATGAAGACGACGATGAAGACGACGATGAAGACGACGATGAAGACGACGATGAAGACGACGA 717

QY 2450 agaaga 2455

DB 718 TGAAGA 723

RESULT 27

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BH148582 942 bp DNA linear GSS 27-AUG-2001
ENTR62TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.

BH148582
BH148582.1 GI:15308029
GSS.

Entamoeba histolytica.
Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.

1 (bases 1 to 942)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica

HM1:IMSS sheared DNA library (2001)
Unpublished (2001)

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlouftus@tigr.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library

Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 794.

Location/Qualifiers
1. .942

FEATURES
source

/organism="Entamoeba histolytica"
/strain="HM1:IMSS"

/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolyt
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."

BASE COUNT 399 a 84 c 240 g 219 t

Query Match 5.3%; Score 176; DB 12; Length 942;
Best Local Similarity 55.3%; Pred. No. 1.2e-13;
Matches 363; Conservative 0; Mismatches 290; Indels 3; Gaps 1;

QY 1855 aaagagcagatttcgaattagaagaagcactcagaagctccagaaggaaggaattatgat 1914

DB 80 ATGATGAAGATGATGAATTTCCAATTAGACGAAGAAGATGATGAAGAAGATGAAGAAG 139

```
Qy 1915 tagaaggagaaggaaacccaacggaagaagccaaagaggaaggaagcgaacacagaagag 1974
Dy 140 ATGAAGAAGATGAAGAAGATCAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAG 199
Qy 1975 aagtgcagagaagaattagagcgaacccagaggaagcagattcgaattagaagaaccaa 2034
Dy 200 AAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATG 259
Qy 2035 caggagaagaagtgaagaacccctagaggcgaagaagcgaagaagcgaagaaggaagaagt 2094
Dy 260 AAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAG 319
Qy 2095 aagagggtacctgcagaagtagaagaagtgggaagggtacctgcagaagtagaagaagtgg 2154
Dy 320 ACGAAGAAGACGAAGAAGATCAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAG 379
Qy 2155 aagagggtaccagaagaagtagaagaaggtaccgcgagaagtagaagaagtagaagaaggtac 2214
Dy 380 AAGACGAGAAGACGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAG 439
Qy 2215 cagaagaagtgggaaggtaccagaagaagtggaaggtaccagaagaagtggaaggtggaag 2274
Dy 440 AATATGATTTAGAGAGAGAGAGATGATGATGCTGTTCGCCAGCTTTTGTTCGTCG 499
Qy 2275 taccagaagaagtgggaagaagtgaagaagtagaagaagtagaaggtaccagcgtagtag 2334
Dy 500 AATTGATGATGAGGACCTCTTGACCTGTGATGAAGGCTTTGATGATGATGATGATGATGAT 559
Qy 2335 aagtagaagttaccagcgtag---tagaagaagaggtgccagaagaagtagaagaagaag 2391
Dy 560 ATCTTGATGATGATGATGATGATGATTTAGAAAGACGACAGACTTTTGAAGACGAGAAAGATG 619
Qy 2392 aagaagaggaagaccagtagaagaagaagtgattacattagtagtaataccatcggaag 2451
Dy 620 AAGAAGATGATGATCAAGAAGACGAGATGATGATTAATCAAGATGGTCAAGAAGAG 679
Qy 2452 aagatatacattagacaacccaagaagaagcgaattaggctcgtggaattttatct 2507
Dy 680 CAAAGTCAGAGGTTTCATCCATTAGATAAAATTGAATCAGCAGACGACCTTTCTTCT 735
```

```
RESULT 28
AZ550718/c
LOCUS
DEFINITION ENTEM36TF Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000
genomic, DNA sequence.
ACCESSION AZ550718
VERSION AZ550718.1 GI:11176019
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 879)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 25
High quality sequence stop: 801.
Location/Qualifiers
1..879
/organism="Entamoeba histolytica"
```

```
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

BASE COUNT 193 a 219 c 72 g 395 t
ORIGIN

Query Match 5.3%; Score 175.8; DB 12; Length 879;
Best Local Similarity 55.5%; Pred. No. 1.3e-13;
Matches 339; Conservative 0; Mismatches 272; Indels 0; Gaps 0;

Qy 1822 aagaagtcgaacggaagaattgatgcacactccaagaggacgatttcgaattagaagaa 1881
Dy 681 ACGTAGAGCTTTTGAAGAAGAAATAATGATGAAGATGATGAATTCCAATTAGACGAAG 622
Qy 1882 ctgcagaagctccagagaaggaattagattagtaggaaggaaggaaggaaggaag 1941
Dy 621 AAGAAGATGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAG 562
Qy 1942 aagaagcaagagagagcgaacacagaaggaagtgccagaagaagaattagagca 2001
Dy 561 AAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATG 502
Qy 2002 ctccagagagcagatttcgaattagaagacccaagaggaagaagtagagaagaaaccctag 2061
Dy 501 AAGAAGATGAAGAAGACGACGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAG 442
Qy 2062 agggcagaagaactgcagaaggaagaagtggaagaggtacctgcagaagtagaagaag 2121
Dy 441 ATGAAGAAGATGAAGAAGATGAAGAAGACGAAGAAGACGAAGAAGATGAAGAAGATGAAG 382
Qy 2122 tggaagaggtacctgcagaagtagaagaagtggaagaggtaccagaagaagtagaagag 2181
Dy 381 AAGATGAAGAAGATGAAGAAGATGAAGAAGACGAAGAAGACGAAGAAGACGAAGAAGACG 322
Qy 2182 taccgcagaagtagaagaagtggaagaggtaccagaagaagtggaagaggtaccagaag 2241
Dy 321 AAGAAGACGAAGAAGACGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAG 262
Qy 2242 aagtgcgaaggtaccagaagaagtggaagaggtaccagaagaagtggaagaggtggaag 2301
Dy 261 ATGCTGTTCGCGAGCTCTTTGCTGCTGAATTTGATGATGATGAGGACCTTCTTGACTCTG 202
Qy 2302 aagtagaagaagtagaggtaccagcggttagtagaagtagaagtagaccagcggttagtagaag 2361
Dy 201 AAGCTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 142
Qy 2362 aagaggtgccagaagaagtagaagaagaagaagaaggaaggaaggaaggaaggaaggaag 2421
Dy 141 CAGACTTTTGAAGACGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATG 82
Qy 2422 atgtattacaa 2432
Dy 81 AATTAACCTCAA 71

RESULT 29
BM170146
LOCUS
DEFINITION EST572669 PyBS Plasmodium yoelii yoelii cDNA clone PYCB91 5' end,
BM170146 832 bp mRNA linear EST 04-DEC-2001

QY	2109	gaagtagaagaagtggaagaggtacctctgcagaaagtagaagaagtggaagaggtaccagaa	2168
Db <td>507</td> <td>GAATCACCAGCAGAAAGCTGAAGTACCAATAGAAGCTGAATCACCAGTGAATCA</td> <td>566</td>	507	GAATCACCAGCAGAAAGCTGAAGTACCAATAGAAGCTGAATCACCAGTGAATCA	566
QY <td>2169</td> <td>gaagtagaagaaggtaccgcgcagagtagaagaagtggaagaggtaccagaagaagtgga</td> <td>2228</td>	2169	gaagtagaagaaggtaccgcgcagagtagaagaagtggaagaggtaccagaagaagtgga	2228
Db <td>567</td> <td>CCAGTAGAAGTAGAAGCAGCAGTAGAAGTAGAGGTGCCAGTAGAAGCTGAAGTAGCTGA</td> <td>526</td>	567	CCAGTAGAAGTAGAAGCAGCAGTAGAAGTAGAGGTGCCAGTAGAAGCTGAAGTAGCTGA	526
QY <td>2229</td> <td>gaggtaccagaagaagtggaagaggtaccagaagaagtggaagaggtaccagaagaagtg</td> <td>2288</td>	2229	gaggtaccagaagaagtggaagaggtaccagaagaagtggaagaggtaccagaagaagtg	2288
Db <td>627</td> <td>GAATGAAGCAGCAGCAGTGAATCAGAAAGCAGTAGTAAGAGTCTGAAGTACCAATAGAAGCT</td> <td>686</td>	627	GAATGAAGCAGCAGCAGTGAATCAGAAAGCAGTAGTAAGAGTCTGAAGTACCAATAGAAGCT	686
QY <td>2289</td> <td>gaagaagtggaagaagtagaagaagtagaaggtaccagcgttagtagaagtagaagtagca</td> <td>2348</td>	2289	gaagaagtggaagaagtagaagaagtagaaggtaccagcgttagtagaagtagaagtagca	2348
Db <td>687</td> <td>GAATCACCAGTAGAAGTAGAAGCACCA--GTGAAGCTGAATCACCAGTAGAAGCTGA</td> <td>743</td>	687	GAATCACCAGTAGAAGTAGAAGCACCA--GTGAAGCTGAATCACCAGTAGAAGCTGA	743
QY <td>2349</td> <td>gcggtagtagaagaagaaggtgcgcagaagaagtagaagaagaagaaggaagaacaa</td> <td>2408</td>	2349	gcggtagtagaagaagaaggtgcgcagaagaagtagaagaagaagaaggaagaacaa	2408
Db <td>744</td> <td>TCACCAGTAGAAGCTGAATCACCAGTAGAAGCTGAATCACCAGTAGAAGCTGAATCACC</td> <td>803</td>	744	TCACCAGTAGAAGCTGAATCACCAGTAGAAGCTGAATCACCAGTAGAAGCTGAATCACC	803
QY <td>2409</td> <td>gtagaagaagaagtagtattaccaa</td> <td>2432</td>	2409	gtagaagaagaagtagtattaccaa	2432
Db <td>804</td> <td>GTAGAAGTAGAAGCACCAGTAGAA</td> <td>827</td>	804	GTAGAAGTAGAAGCACCAGTAGAA	827

RESULT 30

AZ972907/c

LOCUS

2M0246M24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0246M24 R, DNA sequence.

ACCESSION

AZ972907

VERSION

AZ972907.1

KEYWORDS

GI:13844134

SOURCE

GSS.

ORGANISM

house mouse.

REFERENCE

Mus musculus

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 718)

TITLE

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.

JOURNAL

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

COMMENT

Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0246 row: M column: 24
Seq primer: CACACGAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 718.
Location/Qualifiers
1. .718
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0246M24"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource
http://www.jax.org/resources/documents/dnares/. The DNA was hydrodynamically sheared by repeated passage through a

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 10:11:49 ; Search time 72.28 Seconds
(without alignments)
11340.329 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gaattccgtaagtaacaa.....tcttataaaataaataattc 3337

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2.6/prodata/1/lna/5A_COMB.seq:*
2: /cgn2.6/prodata/1/lna/5B_COMB.seq:*
3: /cgn2.6/prodata/1/lna/6A_COMB.seq:*
4: /cgn2.6/prodata/1/lna/6B_COMB.seq:*
5: /cgn2.6/prodata/1/lna/PCTUS_COMB.seq:*
6: /cgn2.6/prodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3337	100.0	3337	1 US-08-072-610-1	Sequence 1, Appli
2	3337	100.0	3337	2 US-08-719-822B-1	Sequence 1, Appli
3	3337	100.0	3337	4 US-09-092-458-1	Sequence 1, Appli
4	199	6.0	3489	2 US-08-728-323A-1	Sequence 1, Appli
5	199	6.0	32207	2 US-08-770-379-20	Sequence 20, Appli
6	199	6.0	32207	4 US-08-757-669A-20	Sequence 20, Appli
7	199	6.0	32207	4 US-09-230-371A-20	Sequence 20, Appli
8	175	5.2	3095	6 5231168-1	Patent No. 5231168
9	170.2	5.1	5361	4 US-08-973-462-2	Sequence 2, Appli
10	170.2	5.1	6152	4 US-08-973-462-1	Sequence 1, Appli
11	167.2	5.0	2580	3 US-09-050-863-2	Sequence 2, Appli
12	167.2	5.0	2580	4 US-09-359-081-2	Sequence 2, Appli
13	167.2	5.0	5452	2 US-09-130-114-1	Sequence 1, Appli
14	167.2	5.0	9800	4 US-08-910-647-1	Sequence 1, Appli
15	167.2	5.0	10596	1 US-07-884-811-15	Sequence 15, Appli
16	167.2	5.0	10596	1 US-07-885-971-15	Sequence 15, Appli
17	167.2	5.0	10596	1 US-08-087-783A-15	Sequence 15, Appli
18	167.2	5.0	10596	1 US-08-194-088B-15	Sequence 15, Appli
19	167.2	5.0	10596	1 US-08-194-087-15	Sequence 15, Appli
20	167.2	5.0	10596	5 PCT-US93-04648-15	Sequence 15, Appli
21	147.2	4.4	9551	1 US-08-056-200-93	Sequence 93, Appli
22	147.2	4.4	9551	2 US-08-800-644-93	Sequence 93, Appli
23	145	4.3	696	4 US-09-461-697-193	Sequence 193, App
24	145	4.3	699	4 US-09-461-697-191	Sequence 191, App
25	145	4.3	717	4 US-09-461-697-189	Sequence 189, App
26	145	4.3	774	4 US-09-461-697-187	Sequence 187, App
27	145	4.3	819	4 US-09-461-697-185	Sequence 185, App

28	145	4.3	1669	4 US-09-461-697-184	Sequence 184, App
29	141.8	4.2	7218	1 US-08-232-463-14	Sequence 14, Appli
30	138	4.1	9636	1 US-08-323-170B-1	Sequence 1, Appli
31	138	4.1	9636	4 US-08-954-441-1	Sequence 1, Appli
32	133.4	4.0	16442	3 US-08-781-891-208	Sequence 208, App
33	127.2	3.8	543	6 5273901-6	Patent No. 5273901
34	124	3.7	1276	4 US-09-177-325-2	Sequence 2, Appli
35	124	3.7	1276	4 US-09-411-812A-2	Sequence 2, Appli
36	124	3.7	1276	4 US-09-590-113-2	Sequence 2, Appli
37	124	3.7	3279	4 US-08-446-137B-1	Sequence 1, Appli
38	117.4	3.5	1891	4 US-08-973-462-3	Sequence 3, Appli
39	116.6	3.5	533	6 5482709-5	Patent No. 5482709
40	115.6	3.5	3211	2 US-08-574-959A-8	Sequence 8, Appli
41	115.6	3.5	3211	4 US-09-357-014-8	Sequence 6, Appli
42	115.6	3.5	3901	2 US-08-574-959A-6	Sequence 6, Appli
43	115.6	3.5	3901	4 US-09-357-014-6	Sequence 6, Appli
44	110.8	3.3	5433	3 US-08-929-329-1	Sequence 1, Appli
45	100.6	3.0	2129	4 US-08-975-762-39	Sequence 39, Appli

ALIGNMENTS

RESULT 1
US-08-072-610-1
; Sequence 1, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,610
; FILING DATE: 19930602
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/07686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pWMB3.3.1
US-08-072-610-1

Query Match 100.0% Score 3337; DB 1; Length 3337;

Best Local Similarity 100.0%; Pred. No. 0;				
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	gaattccggttaaaagtaacaaactatgttctcgtatctatataacaccttaattttatc	60	
Db	1	GAATTCGGGTAAAGTAACAACTATGGTTTCGTATCTATATATAACCTTACTAATTTTATC	60	
QY	61	ttttgctttcttttaattcatcttcaacagtaagataaaaaataatctataaaactgc	120	
Db	61	TTTTTGCTTTTCTTTAATTCATGCTTCAACAGTAAGATAAATAATATATATAAACTGC	120	
QY	121	tatatatacatatataatcctaagtggtcatttgtgaattgcgatcatcttaatttacgta	180	
Db	121	TATATATACATATATATCAATAGTGGCATTTGTGAATTCGGATCATTTAAATTTACGTA	180	
QY	181	aaacaataattgaaaaaatt	240	
Db	181	AAACAATAATTGAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTGTCTACAGACGATTTAG	240	
QY	241	aattggaaaatgctctctgatgatgtttagaggtgagggatccctcaaacgacggtttag	300	
Db	241	AATTGGAAAATGCTTCTGATGATGTTGTACAGGTGGAGGATCTTCAAAACGACGGTTTAG	300	
QY	301	aattagaagaggaaaatttttgatgagaattccaggtgatgatgaactcttttagtgcta	360	
Db	301	AATTAGAAGAGGAAAATTTTGTAGAGAAATTCAGGTGATGATGAACCTCTTTTAGATGCTA	360	
QY	361	ccccggaagatgactttgccttaacagatttgccaatggaagcgtagaggaagtcacag	420	
Db	361	CCCCGGAAGATGACTTTGCCCTTAACAGATTTGCCAATTTGAAGCATGAGGAAGTCAACG	420	
QY	421	aaacgttagatgaggtgaaatcattaggagaggtttccactgaagatatgaaacagaag	480	
Db	421	AAACGTTTAGATGGAGGTGAATCATTTAGGAGAGGTTTTCCACTGAAGATATGGAACAGAA	480	
QY	481	atggctcaacagatgatacagaacagaagaagaaactcctggtgatatagaaggagaag	540	
Db	481	ATGGCTCAACAGATGATACCGGAACAGAAGAAGGACTTACCCTGGTGATATGGAAGGAGA	540	
QY	541	aagaagctggcgtatgtaagcagcggggaagaagctggtgatttggagcaggggaagaaa	600	
Db	541	AAGAAGCTGGGTATATGGAAGCAGGGGAAGAAGCTGGTGATTTGGAAGCAGGGGAGAAA	600	
QY	601	ctggcgtattggaagcagggggaagaaactggcgtatttggagcaggggaagagctgggt	660	
Db	601	CTGGCGATTTTGGAGCAGGGGAAGAACTGGCGATTTTGGAGCAGGGGGAAGAGCTGGTG	660	
QY	661	atttgaagcaggggaagaaactggcgatttgaagcaggggaagaaactggagatgcgg	720	
Db	661	ATTTGGAGCAGGGGAAGAACTGGCGATTTTGGAGCAGGGGGAAGAACTGGAGATGCGG	720	
QY	721	aaactgaagagagcaactggagatgcggaactgaaactgaaactggagcaactgtgtatgtag	780	
Db	721	AAACTGAAGAAGAGGACCACTGGAGATGCGCAAACTGAAATTTGGAAGCAACTGTGTATGTAG	780	
QY	781	acacagaagatagttcagctgataggagcagaaaaagtaacatgttctctgctcaagaaaaatg	840	
Db	781	ACACAGAAGATAGTTTCACTAGCTGATGGAGCAGAAAAAGTACATGTTCCTGCTCAAGAAAA	840	
QY	841	tacaacctccgatagtaatgatgccctcttggaaagtatttggataaagatatattt	900	
Db	841	TACAACTCGCGATAGTAATGATGCCCTCTTTTGGAAAGTATTTTGGATTAAGATATATATT	900	
QY	901	ttgatcataatgaatttcagccactatttcgaacaaaatttgtggcggttactgctaaac	960	
Db	901	TTGATCATATTAAGATTTCAGGCCACTATTCGAACAAATTTGTGGCGGTACTGCTTAAC	960	
QY	961	atgttcagggaagaatttcgaatgaacactgtaccattaccagtggcagaagagcccg	1020	
Db	961	ATGTTACGGGACAGAAATTCGAATGAACCTGTACCATTACCAGTGGCAGAAGAGCCCG	1020	
QY	1021	cgcaagtaccagcggaagattagatgccactccagaggtgatcttcgattagattta	1080	

Db	1021	CGCAAGTACCAGCGGAAGAAATTAGATGCCACTCCAGAGGATGACTTCGCATTAGATGTTA	1080	
QY	1081	cagaactctccaggaagtagaattagttatgaagcaggaactgaagaagaatcaa	1140	
Db	1081	CAGAATCTCCGAGGAAGTAGAATTAATTAGTGAAGGCACTGAAGAGATCAA	1140	
QY	1141	cggaagtgggaccaacggaaggaagcccaacccgaagaatttagtgcctctccagagatg	1200	
Db	1141	CGGAAGTGGACCACACGGAAGAGGACCAACCCGAAGAATTAGATGCCACTCCAGAGGATG	1200	
QY	1201	gatttcgacttagacgaactgcagaggaagagaaaaagagaaactgtagagggagagaaaa	1260	
Db	1201	GATTTTCGCATTAGACGAACCTGCAGAGGAGAGAAACAGAAACGTAGAGGGAGAGAAA	1260	
QY	1261	cagaagaagctgcagaaggaagatcagaagaactccagaaggaagagaagaagttag	1320	
Db	1261	CAGAAGAAGCTGCAGAGGAGAAAGTATCAGAAGAACTCCAGAGGAGGAAGAGAGTTAG	1320	
QY	1321	aggcaactccagaggatatttcgcattagatggaactacattagaagaacccaagaaa	1380	
Db	1321	AGGCAACTCCAGAGGATGATTTGCGATTAGATGGAACCTACATTAGAGAAACCGAAGAA	1380	
QY	1381	ctgcagaagagaagaaccgttagaggaagaacacccgttagagcagaagaacccgtag	1440	
Db	1381	CTGCAGAAGGAGAGAAACCCGTAGAGGGAGAGAAACCCGTAGAGGGAGAGAAACCCGTAG	1440	
QY	1441	agggaagaagctgcagaaggaagaagttagagcgaactccagagagatgacttcc	1500	
Db	1441	AGSGAGAAGAAGCTGCAGAAAGGAGAGAGATTAGAGGCAACTCCAGAGGATGACTTCC	1500	
QY	1501	aattagaagaaccatcagagaaggaaggggaaggaaggaaggaaggaaggaaggaag	1560	
Db	1501	AATTAGAAGAACCATCAGGAGAGGAGAGGGGAAGGAGAGGAAGGGGAGGAGGAAG	1560	
QY	1561	gagaagcgttagtagcagtcagtagtgccgaacccgttagaagtagtgcctcctgc	1620	
Db	1561	GAGAAGCGTTAGTAGCAGTGCCAGTAGTGCCCAAGCCGTAGAGAGTAGTACTCTCTGCTC	1620	
QY	1621	agcctgtcaaaccaactgctcccaacgcgagatgaaactttattcgttgatatttag	1680	
Db	1621	AGCCTGTCAAAACCAATGCTCGCTCCAACGCGCAGATGAAACTTTATTCTGTTGATATTAG	1680	
QY	1681	ataacgatttaacgtatgcagacattacattcctttgagccattatttaacaatactca	1740	
Db	1681	ATAACGATTTAACGATGTCAGACATTACATCTCTTTGAGCCATTATTATAACAATCTCTCA	1740	
QY	1741	aggatcctgatgcagagagcgtgtacacagtaacatcaaggaagcaccctgttacaagtac	1800	
Db	1741	AGGATCCTGATGCAGGAGAGGCTGTAACTACCATCAAGGAAGCAACCTGTACAAAGTAC	1800	
QY	1801	cagtggcagtagggccgcgcaagaagtgcacaacggaagaattgatgcaactccaagagg	1860	
Db	1801	CAGTGGCAGTAGGGCCCGCCGAAGAGTGCCAACGGGAAGAAATTTGATGCAACTCCAAGAGG	1860	
QY	1861	acgatttcgaatttagaaggaactgcagaagctccagaggaaggaagatttagttagaag	1920	
Db	1861	ACGATTTCGAATTAGAAAGGAACCTGCAGAGCTCCAGAGGAAGGAGAAATTAGTATTAGAAG	1920	
QY	1921	gagaaggaagaacacgcggaagaagagcgaagagaagagagccacaagaaggagaagtcg	1980	
Db	1921	GAGAAGGAGNACCACGGAAGAAGAGCCCAAGAGAGAGAGAGCCCAACAGAAGGAGAGTGC	1980	
QY	1981	cagaagaagaatttagaggccaactccagagagcatttcgaattagaagaaccaacaggag	2040	
Db	1981	CAGAAGAAGAAATTAGAGGCAACTCCAGAGGACGATTTCGAATTTAGAAGAACAACAGGAG	2040	
QY	2041	aagaagtagaagaaaacgtagagggcgaagaaactcagaaggaaggaagaagtggaagagg	2100	
Db	2041	AAGAAGTAGAAGAAAACCGTAGAGGGCAAGAAACTGCAGAAGGAGAGAAAGTGGAAAGAGG	2100	
QY	2101	tacctcagaagtagaagtggaaggtacctcagaagtagaagaagtggaagagg	2160	
Db	2101	TACCTGCAGAAAGTAGAAGAAAGTGAAGAGGTACTCTCAGAAAGTAGAAGAACTGGAAGAGG	2160	

QY 2161 taccagaagaagtagaagaggtaccgcgacagaagtagaagaagtggaagaggtaccagaag 2220
Db 2161 TACCAGAAGAAGTAGAAGAGGTACCGCAGACACTAGAGAAGTGGGAAGAGGTACCAGAAG 2220
QY 2221 aagtggaagaggtaccagaagaagtggaagaggtaccagaagaagtggaagaggtaccag 2280
Db 2221 AAGTGAAGAGGTPACCAAGAGAAGTGGGAAGAGGTACCAGAAGAAGTGGGAAGAGGTACCAG 2280
QY 2281 aagaagtggaagaagtggaagaagtagaagaagtagaaggtaccagcggtagtagaagtag 2340
Db 2281 AAGAAGTGGGAAGAAGTGGGAAGAAGTGAAGAAGTGAAGAGTACCAGCGGTAGTAGAAGTAG 2340
QY 2341 aagtcacagcggtagtagaagaagaggtgccagaagaagtagaagaagtagaagaagagag 2400
Db 2341 AAGTACCAGCGGTAGTAGAAGAAGAGGTGCCAGAAGAAGTGAAGAAGAGAGAGAGG 2400
QY 2401 aagaaccagtagaggaagaagtagattacaattagtaaccatcggaagaagatatc 2460
Db 2401 AAGAACCAGTAGAGGAAGAAGTGTATTACAATTAGTAATACCATCGGAAGAAGATATAC 2460
QY 2461 aattagacaaacaaagaagaacgaattaggctctggaatttttatctatcatcgacatgc 2520
Db 2461 AATTAGACAACCAAGAAAGAGCAATTAGGCTCTGGAATTTTATCTATCATCGACATGC 2520
QY 2521 actcacaagacgttccaaaggaatttatggaagaagaagaactgcagtgatccat 2580
Db 2521 ACTTACCAGACGTTCCANAGGANTTATGGAAGAAGAGAAACTGCAGTGTATCCAT 2580
QY 2581 tgaaccagaagattttgcaagaagaagattcacaaattcagaaatggctcacatctc 2640
Db 2581 TGAACCAAGAAGATTTTGAAGGAAGATTCACAATCTCAGAAATGGCTCACATTCATTC 2640
QY 2641 aagccctaagacgactgggaacgattagaagtgagcttaataaggtagagaagaat 2700
Db 2641 AAGCCCTAGAGCGGACTGGGAACGATTAGAAGTGAGCTTAAATAGGCTAGAGAAGAT 2700
QY 2701 ggaatgaaacaaagaataaagaatggctggctggctggctggctggctggctggctggct 2760
Db 2701 GGATGGAACAAAGAAATAAAGATGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 2760
QY 2761 cagaatatagtcataattcaacaaaggaaagacccagctggctggaaacacgagat 2820
Db 2761 CAGAAATATAGTCAAAATTTCAACAAAAAGAAAGGACCCAGCTGGTTTGAGAAAAACGAGAT 2820
QY 2821 ggaagcagcagaagtagaagaatggtttaaagcagaagtcacaaatcccaattgattcac 2880
Db 2821 GGAGCCAGCAGAANTGGAAAAATGGTTTAAAGCAGAAGTCAANTCCCAATTTGATTTCAC 2880
QY 2881 acttgaaaaatggatgaacgacactcattcccaatttttaaaattcttgtaaaagata 2940
Db 2881 ACTTGAAAAATGGATGAACGACACTCATTCCCAATTTATTTAAAAATCTTGTAAGAGATA 2940
QY 2941 tgtcacaaattgaaacacagaacaaacagaatggttaatactactggaacaaagaacg 3000
Db 2941 TGTCACAATTTGAAAAACAAGAAACCAAGAAATGGTTAATGAATCACTCGAAAAAGAACG 3000
QY 3001 aacggggttatggtctgaatcatttgaaagttagccacatcaaaattattaaatgtg 3060
Db 3001 AACGGGGTTATGGTCTGTAATCTTTGAAGTTATGACCACATCAAAATTTATTAATTTGG 3060
QY 3061 ctaagagtcgagaatggtaccgtgccaatctctaataataagagaagaagaagaactca 3120
Db 3061 CTAAAGAGTCAGAATGGTACCGTGCCAATCTTAATAATAAGAGAAGAAAGAACTCA 3120
QY 3121 tgaataatgggtttccctaaagaacgaataatttaggacaaagaatggaacaaatggact 3180
Db 3121 TGAATATGGGTTTCCCTAAAGAAAGCAATATTTAGGACAAAGAAATGGAATAATGGACT 3180
QY 3181 cattgaaaaaagttaaatttttggttcattcaattcaatggtgacacattttctgaaaa 3240
Db 3181 CATTGAAAAAAGTTAAATTTTGGTTCAATTTCAATGTTGTAACACATTTTCTCGAAAA 3240

QY 3241 cgctaaccagaagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaag 3300
Db 3241 CGCTAACCAAGCAAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3300
QY 3301 aaagaacagagattctctctataaaataaataattc 3337
Db 3301 AAAAGAAGACAGATTATTCTCTTATATAAATAAATAATTC 3337
RESULT 2
US-08-719-822B-1
; Sequence 1, Application US/08719822B
; Patent No. 5874527
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,822B
; FILING DATE: 09/30/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: PMB3.3.1
US-08-719-822B-1

Query Match 100.0%; Score 3337; DB 2; Length 3337;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaattccggtaaagtaaacactatggttcgtatctatataaaccttactatcttattc 60
Db 1 GAATTCGGTAAAGTAACACTATGTTTCGTATCTATATATAACCTTACTAATTTATC 60
QY 61 ttttgcgtttcttttaatttcatttcacagtaagataaaataatctataaaactgc 120
Db 61 TTTTGTGTTTCTTTAATTCATGCTTCACAGTAAGATAAAATAATCTATAAAACTGC 120
QY 121 tatataacatatatttcacagtggtcatttgatgacgatttgaatttcacgta 180
Db 121 TATATACATATATATTTGTAAGTGGCATTTGTGAATTCGATCATTTAAATTTACGTA 180

[illegible][illegible]

Db 2341 AAGTACCAGCGGTAGTAGAAGAAGAGGTGCCAGAAGAAGTAGAAGAAGAAAGAGAGG 2400
QY 2401 aagaaccagtagaggaagatgtattacaattagtaataccatcggaagaagatatac 2460
Db 2401 AAGAACCGTAGAGGAAGAAGATGTATTACAAATTAGTAATACCATCGGAAGAAGATATAC 2460
QY 2461 aattagacaaaccaaagaagaacgaattaggctctggaattttatctatcatcgacatgc 2520
Db 2461 AATTAGACAAACCAAAGAAGACGAATTAGGCTCTGGAATTTATCTATCATCGACATGC 2520
QY 2521 actaccaagacgttccaaaggaaattattggaagaagaagaactgcaagtgtatccat 2580
Db 2521 ACTACCAAGACGTTCCAAAGGAATTATGGAAGAAGAAGAACTGCAAGTGTATCCAT 2580
QY 2581 tgaaccagaagatttgcgaagaagattcacaattcacagaaatggctcacattcattc 2640
Db 2581 TGAACCCAGAAGATTTGCAACGAAGATTCACAATCTACAGAAATGGCTCACATTCATTC 2640
QY 2641 aaggcctagaagcgactgggaacgattagaagtgagcttaataaggctagagaagat 2700
Db 2641 AAGGCCTAGAAGCGACTGGGAACGATTAGAAGTGAGCTTAATAAGGCTAGAGAAAGAT 2700
QY 2701 ggaatgaacaaagaataaagaatggctggctggcttgcgttaattgaaataaaatggt 2760
Db 2701 GGATGGAACAAAGAAATAAAGAAATGGCTGGCTGGCTTTAAATGAAATFAAATGGT 2760
QY 2761 cagaaatagtcataatttcaacaaaggaagaccagctggtttgagaacacgagagt 2820
Db 2761 CAGAATATAGTCAAAATTCACAAAGGAAGAACGCCAGCTGGTTGCAGAAACGAGAGT 2820
QY 2821 ggagcagcagagaatggaaataatgggttaagcagaagtcacaaatccccaaattgattcac 2880
Db 2821 GGAGCGACGAGAAATGMAAAATGGTTAAAGCAGAAGTCAAAATCCCAAATTTGATTAC 2880
QY 2881 acttgaataaatggatgaacgacacattccatttattttaaattcttgtgaaagata. 2940
Db 2881 ACTTGAATAAATGGATGAACGACATCTCCAAATTTATTAAATTCCTTTGTGAAAGATA 2940
QY 2941 tgtcacaatttgaacaaacaaacaaagaatgggttaatgaatcactgggaacaaagacg 3000
Db 2941 TGTCACAATTTGAAACAAACAAACAAACGAATGGTTAATGAATCACTGCAAAACGAACG 3000
QY 3001 aacggggttagtcttgatcatttgaaagttatgaccacatcaaaattttaaattgtg 3060
Db 3001 AACGGGTTATGGTTCTGAATCAATTTGAAGTTATGACCACATCAAAATTTAATATGG 3060
QY 3061 ctaagatcgagaatggtaccgtgccaatcctaataataagagaagaagaaactca 3120
Db 3061 CTAAAGATCGAGAATGGTACCGTGCCAAATCCTAATAATAAGAGAAGAGAACTCA 3120
QY 3121 tgaatgggttctcctaaagaaacgaaatatttaggacaaagaatgggaataatggact 3180
Db 3121 TGAATGGTTCTCCTTAAAGAAACGAATATTTAGGACAAACAAATGGAAATAATGGACT 3180
QY 3181 cattgaaaaaagctaaatttttggtttcaattcgaatggtacacattttcttgaaaa 3240
Db 3181 CATTTGAAAAAAGTTAAATTTTGGTTTCAATTCATGTTGACAAATTTCTTGGAAAA 3240
QY 3241 cgcttaaccaagaagaatggaatcatttgaatgaaataaaagtgtgaattatagaa 3300
Db 3241 CGCCTAACCCANGGAAGAAATGGAATCAATTTGTTAATGAAATAAAGTTTGAATATAGAA 3300
QY 3301 aaagaacagattattctcttaaaaaataaattc 3337
Db 3301 AAAAAGACAGATTATCTCTTATAAAATAAATAATTC 3337

RESULT 3

US-09-092-458-1
; Sequence 1, Application US/09092458
; Patent No. 6231861
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John

; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,458
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/719,821
; FILING DATE: 09/30/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: PVMB3.3.1
; US-09-092-458-1

Query Match 100.0%; Score 3337; DB 4; Length 3337;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 gaattccgggttaaagtaacaactatggttcgtatctctatataacaccttactaatctac 60
Db 1 GAATTCGGGTAAAGTAACAACACTATGGTTTCGTATCTATATATACCTTACTAATTTATC 60
QY 61 ttttgcctttctttaaattcatcttcacagtaagataaaaaatactataaaaaactgc 120
Db 61 TTTTGTCTTTCTTTAAATTCATGCTTCAACAGTAAGATAAAAAATAATCTATAAAAACTGC 120
QY 121 tatatatacatatataatcagaagtgccatttggaattgcgatacatttaaatctagta 180
Db 121 TATATATACATATATATATATAAGTGGCATTTTGGAATTCGGATCATTTAAATTTACGTA 180
QY 181 aaacaataattgaaaaaatt 240
Db 181 AAAACAATATGAAAAAATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCAT 240
QY 241 aattgaaaaatgcttcttgatgattgttagagggtggaggatccttcaaacgacggttag 300
Db 241 AATTGGAAAAATGCTTCTGTATGATGCTTTGTAGAGGTGGAGGATCCTTCAACGACCGTTAG 300
QY 301 aattagaagaggaataattttgtagagaatccagggtgatgatgaacatcttttagatgcta 360
Db 301 AATTAGAAGAGGAAAAATTTTTCATGAGAATTCAGGTGATGATGATGATGATGATGATGATGAT 360

QY 361 cccccgaagatgacttgccttaacagattgccaattgaagcagatgaggaagtcaacg 420
Db 361 CCCCGAAGATGACTTTGCGCTTAACAGATTGCCAATTGAAGACGATGAGGAAGTCAACG 420
QY 421 aaacgttagatggaggtgaatcattagagagaggtttccactgaagatatggaacagaag 480
Db 421 AAACGTTAGATGGAGGTGAATCATTTAGAGAGGTTTCCACTGGAAGATATGGAACAGAAG 480
QY 481 atgctcaacagatgatacagaaacagaagaagactacacctggtgatatggaagagaag 540
Db 481 ATGGCTCAACAGATGATACGGAACACAGAGAGGACTACCTGGTGATATGGAAGCAGAAG 540
QY 541 aagaagctggcgatatgaaagcaggggaagaagctggtgatttggagcagggggaagaaa 600
Db 541 AAGAAGCTGGCGATATGAAGCAGGGGAAGAAGCTGGTGATTTGGAACGAGGGGAAGAAA 600
QY 601 ctggcgatttggaaagcaggggaagaactggcgatttggaaagcaggggaagaagctggtg 660
Db 601 CTGGCGATTTGGAAGCAGGGGAAGAAAACCTGGCGATTTTGGGAAGCAGGGGAAGAACTGGTG 660
QY 661 atttggagcaggggaagaactggcgatttggagcaggggaagaactggagatcgcg 720
Db 661 ATTTGGGAAGCAGGGGAAGAACTGGCGATTTTGGGAAGCAGGGGAAGAACTGGAGATCGGG 720
QY 721 aaactgaagaagggcaactggagatcggaactggaactgaaactggagcaactgtgtatgag 780
Db 721 AAACCTGAAGAAGGAGCAACTGGAGATCGGGAACCTGGAATGGAGCAACTGTGTATCTAG 780
QY 781 acacagaagatagttcagctgtaggagcagagaagaagtacatgttccctgctcaagaagaatg 840
Db 781 ACACAGAAGATAGTTCAGCTGATGGAGCAGCAAAAAGTACATGTTCCTGCTCAAGAAAATG 840
QY 841 tacaacctggcgatagtaactgacctcttggaaatttggataaagataatatt 900
Db 841 TACAACCTGGCGATAGTAAGTATGCCCTCTTTGGAAATTTTGGATAAAGATATAATTT 900
QY 901 ttgatcatattaaagatttcagccactatttcgacaacttggcggggtactgctaaac 960
Db 901 TTGATCATATTAAAGATTTCAGGCCACTATTTCGAACAAATTTGTGGCGGGTACTGCTAAAC 960
QY 961 atgttcaggacaagaattccaatgaaacctgtaccattaccagtggcagagaagcccg 1020
Db 961 ATGTTACGGGCACAAATTCGCAATGAACCTGTACCATTACCAGTGGCAGAGAGCCCG 1020
QY 1021 cgcaagtaccagcggaagaattagatgccactccagaggtgacttcgcattagatgta 1080
Db 1021 CGCAAGTACCAGCGGAAGATTAGATGCCACTCCAGAGGATGACTTCGCATTTAGATGTTA 1080
QY 1081 cagaatctcccgagagtagaattagattagatgaagggcaactgaagaagaatcaa 1140
Db 1081 CAGAATCTCCGAGGAAGTAGAATTAGTATTAGATGAAGGCAACTGAAGAAGAATCAA 1140
QY 1141 cggagttggaccacgaagaagcccaaccgaagaattagatgccactccagagatg 1200
Db 1141 CGGAAGTGGGACCAACCGAAGAGACCAACCGAAGAATTAGATGCCACTCCAGAGGATG 1200
QY 1201 gatttcgattagacgaactgcagaagggagaacagaagaactgtagaggggaagaaa 1260
Db 1201 GATTTGCGATTTAGCAAACTGCAAGAGGAGAAACAGAGAACCTGAGGGGAGAAGAAA 1260
QY 1261 cagaagaagctgcagaagggagaagatcagaagaactccagaagaagcagaagaagtag 1320
Db 1261 CAGAAGAAGCTGCAAGAGGAGAACTATCAGAAGAAACTCCAGAAGAGGAGAAGAGTTAG 1320
QY 1321 agccaactccagagatgatttcgattagatggaactacattagaagaacccaagaaa 1380
Db 1321 AGGCAACTCCAGAGATGATTTCCATTAGNTGGAACTACATTAGAGAARAACCGAAGAAA 1380
QY 1381 ctgcagaaggaagaacccgtagaggagaagaacccgtagagggaagaagaacccgtag 1440
Db 1381 CTGCAGAAGGAGAAGAAACCGTAGAGGGAGAAGAAACCGTAGAGGGGAGAAGAAACCGCTAG 1440

QY 1441 agggagaagaagctgcagaagggagaagaagtttagaggccaactccagagagatgacttcc 1500
Db 1441 AGGAGAGAAGAGTGCAGAAGGAGAGAAGAGTTAGAGGCAACTCCAGAGGATGACTTCC 1500
QY 1501 aattagaagaacatcaggaagaaggaaggaagcagaagaaggaaggaaggaag 1560
Db 1501 AATTAGAAGAACCATCAGGAGAAGGAGGGAAGGAGGAGGAGGGAAGGAGAGAAG 1560
QY 1561 gagaagcgttagtagcagtgccagtagtggccgaaccggtagaagtagtagtactctctc 1620
Db 1561 GAGAAAGCTTTAGTAGCAGTGCCAGTAGTGCCGAACCCGTAGAAGTAGTGACTCTCTC 1620
QY 1621 agcctgccaacccaatggtcgtcccaacggcagatgaacttttattcgttgatctctag 1680
Db 1621 AGCCTGTCAAACCAATGTCTCCAACGGCAGATGAACATTTATTCTGTTGATATCTTAG 1680
QY 1681 ataacgatttaacgtatgcagacattacatcttcttgaccattatttaacaacatctca 1740
Db 1681 ATAACGATTTAACGTATGCAGACATTACATCTTTTGAGCCATTATTAAACAATCTCA 1740
QY 1741 aggtacctgtagcaggaagcgtgtaacagtaccatcaagaagcagccctgtacaagtac 1800
Db 1741 AGGATCTGATGCAGGAGAGSCTGTACAGTACCATCAAGGAAGCACTGTACAAGTAC 1800
QY 1801 cagtggcagtagggccgcgaagaagtgcacaacggaagaattgatgcaactccaagagg 1860
Db 1801 CAGTGGCAGTAGGGCCCGCAGAACTGCAGAACTGCACGGAAGATTTGATGCAACTCCAAGAGG 1860
QY 1861 acgatttcgaattagaagaactgcagaactccagaggaaggaattagattagaag 1920
Db 1861 ACGATTCGAATTAGAAGGAACCTGCAGAGCTCCAGAGGAGGAGAATTAGTATTAGAAG 1920
QY 1921 gagaagcagcaacacgcgaagaagccaaagagagagagcaccagaagggagaagtgc 1980
Db 1921 GAGAAGGAGAACCACCAACGSAAGAGAGCCAAAGAGAAGGAGGCCAACAGAAGGAGAAGTGC 1980
QY 1981 cagaagaagaattagaggcaactccagagagcagatttcgaattagaagaaccaacagag 2040
Db 1981 CAGAAGAAGAAATTAGAGGCAACTCCAGAGGACGATTTTCGAATTTAGAAGAACCAACAGAGG 2040
QY 2041 aagaagtagaagaacccgtagagggcgaagaactgcagaaggaagaagtggaagagg 2100
Db 2041 AAGAACTAGAAGAACCCTAGAGGCCGAAGAACTGCAGAAGGAGAGAAGTGGAAAGG 2100
QY 2101 taactgcagaagtagaagaagtggaaggtacctgcagaagtagaagaagtggaagagg 2160
Db 2101 TACCTGCAGAAGTAGAAGAAGTGGAAAGGTACCTGCAGAAGTAGAAGAAGTGGAAAGAGG 2160
QY 2161 taccagaagaagtagaagaggtaccgcagaagtagaagaagtggaagaggtaccagaag 2220
Db 2161 TACCAGAAGAAGTAGAAGAGGTACCCGCAGAAGTAGAAGAAGTGGAAAGGTACCAAGAG 2220
QY 2221 aagtggaaaggttaccagaagaagtggaaagagggtaccagaagaagtggaagaggtaccag 2280
Db 2221 AAGTGGAAAGGTACCAGAGAAGAGTGGAAAGGTACCAGAAAGTGGAAAGGTACCAG 2280
QY 2281 aagaagtygaagaagtggaagaagtagaagaagtagagggtaccagcgggttagtagaagtag 2340
Db 2281 AAGAAGTGGAAAGTAGCAAGAGTAGAAGAAGTAGAGGTACCAAGCGGTAGTAGAAGTAG 2340
QY 2341 aagtaccagcggtagtagaagaaggtgccagaagaagtagaagaaggaagaagaagagg 2400
Db 2341 AAGTACCAGCGGTAGTAGAAGAGAGGTGCCAGAAGAAGTAGAAGAAGAGAAGAGAGG 2400
QY 2401 aagaacccagtagagaagaagatgtattacaattagtaataaccatcggaagaagataac 2460
Db 2401 AAGAACCTAGAGGAAGAGATGTATTACAATTAGTTAGTAAATACCATCGGAAGAAGATATAC 2460
QY 2461 aattagacaacccaagaagaagaattaggctcgtggaatttttctatcatcatcgacatgc 2520
Db 2461 AATTAGACAACCCAAGAAAGACGAATTAGGCTCTGGAATTTTATCTATCATCTGACATGC 2520
QY 2521 actaccaagacggttccaaagggaatttatggaagaagaagaacgctgagtgatccat 2580

Db 20611 CCTCAGGAGCCACACAGCAGGAGGCCACACAGCAGCAGGAGCCACAGCAGGAGGCCCT 20552
Qy 1428 gaagaaacggtagaggagaaagctgcagaagaggagaagaagtagtagggcaactcca 1487
Db 20551 GCAGGCCACACAGCAGGAGCCACAGCAGCAGGAGCCCTGCAGGAGCCACACAGCA 20492
Qy 1488 gaggatgacttccaattagaagaaccatcagggaaggaagggggaaggaaggaagaa 1547
Db 20491 GGAGCCACACAGCAGGAGCCACAGCAGCAGGAGGAGCCACAGCAGCAGCAGCA 20432
Qy 1548 ggggaaggaaggaagcgttagtagcagtagtccagtagtgccgaacggttagaagta 1607
Db 20431 GGAGCCACAGCAGGAGCCACAGCAGCAGGAGCCACAGCAGCAGGAGCCACAGCA 20372
Qy 1608 gtgactcgtcagcgtgtcaaaccaatgtcgtcccaacgagatgaactttattc 1667
Db 20371 GGAGCCACAGCAGGAGCCACAGCAGCAGGAGCCACAGCAGGAGCCACAGCAGCG 20312
Qy 1668 gttgatcttagataacgtttaacgtatgcagaaattacatccttttgagccatttt 1727
Db 20311 GGAGCCACAGCAGGAGCCACAGCAGCAGGAGCCACAGCAGGAGCCACAGCAGCG 20252
Qy 1728 aacaaatcctcaaggtcctgtatgcagagaggtctgaacgtaccatcaaggaagca 1787
Db 20251 GGAGCCACAGCAGGAGCCACAGCAGCAGGAGCCACAGCAGGAGCCACAGCAGCA 20192
Qy 1788 cctgtacaagtaccagtggcagtagggccgcgcaagaagtgcacacgggaagattgat 1847
Db 20191 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCA 20132
Qy 1848 caactccaagagcagcttctgaattagaaggaaactcagaagctccagagggaagaa 1907
Db 20131 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCA 20072
Qy 1908 ttagtattagaaggagaggaacccaacggaagaagcgcaagagagagagagca 1967
Db 20071 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCA 20012
Qy 1968 gaaggagaagtccagaagaattagaggcaactccagaggacgatttcgaattagaa 2027
Db 20011 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCA 19952
Qy 2028 gaaccaagaggaagtagaagaacccgttagaggcggaagaactgcagaaggaagaa 2087
Db 19951 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCA 19892
Qy 2088 gaagtggaaagggtacctgcagaagtagaagaagtggaaagggtacctgcagaagtagaa 2147
Db 19891 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCA 19832
Qy 2148 gaagtggaaagggtaccagaagaagttagaagggtacccgcagaagtagaagaagtggaa 2207
Db 19831 GCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCA 19772
Qy 2208 gaggtagcagaagaagtggaaagggtaccagaagaagtggaaagggtaccagaagaagt 2267
Db 19771 GCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCAGGATGAGCA 19712
Qy 2268 gaagaggtaccagaagaagtggaaagggttagaagaagtagaagggtaccagcg 2327
Db 19711 GGAGGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCA 19652
Qy 2328 gtagtagaagttagaagtagcagcgttagtagaagaagaggtgcagagaagtagaagaa 2387
Db 19651 GTTAGAGGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCA 19592
Qy 2388 gaagaagaagaaggaaccagtagagaagaagtagtattacaattagtaataccatcg 2447
Db 19591 GGAGTTAGAGGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCA 19532
Qy 2448 gaagaagatatacaattag 2466
Db 19531 GCAGGATGAGCAGCAGCAG 19513

RESULT 7

US-09-230-371A-20/c
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230, 371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

Query Match 6.0%; Score 199; DB 4; Length 32207;
Best Local Similarity 46.9%; Pred. No. 4.2e-27;
Matches 685; Conservative 0; Mismatches 770; Indels 4; Gaps 2;

Qy 1011 gaagagccccgcaagtaccagcggaagaattagatgccactccagaggtgacttcgca 1070
Db 20970 GAGGATCAGCAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAG 20911
Qy 1071 ttgatgtttacagaatctcccgaggaagtagaattagattagatgaagaggcaactgaa 1130
Db 20910 GAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAG 20852
Qy 1131 gaagaatcaacgcggaagtgggacccaacggaagaaggacccaagaaattagatgccact 1190
Db 20851 GGAGGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 20792
Qy 1191 ccagaggttgatttcgcattagacgaaactgcagaagggaaggaaggaaggaagaaactgtag 1250
Db 20791 CGAGGAGGATGACGATGATGAGGACCAATGAGGACGAGGAGGATGACGAGGAGGAGGACAA 20732
Qy 1251 ggagaagaacacagaagaagctgcagaaggagaagtagatcagaagaactccagaaggaagaa 1310
Db 20731 GAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 20672
Qy 1311 gaagagttagaggcaactccagaggtgatttcgcattagatggaactacattagaaagaa 1370
Db 20671 ACAGCAGCAGGAGCCACACAGCAGGAGGAGCAGCAGCAGCAGGAGCCACAGCAGGAGGAGCC 20612
Qy 1371 accgaagaactgcagaaggagaagaacccgttagaggag---aagaaacccgttagagga 1427
Db 20611 CCTGCAGGAGCCACACAGCAGGAGGCCACAGCAGCAGGAGCCACAGCAGCAGGAGGCCCT 20552
Qy 1428 gaagaaaccgttagagggaagaagctgcagaaggagaagaagagtagtagaggccaactcca 1487
Db 20551 GCAGGAGCCACACAGCAGGAGGCCACAGCAGCAGGAGGCCCTGCAGGAGCCACACAGCA 20492
Qy 1488 gaggatgacttccaattagaagaaccatcaggaaggaaggaaggaaggaaggaaggaagaa 1547
Db 20491 GGAGCCACACAGCAGGAGCCACAGCAGGAGGAGCCACAGCAGGAGGAGCCACAGCAGCA 20432
Qy 1548 ggggaaggaagaaggaagcgttagtagcagtagtgcagtagtgccgaacccggttagaagta 1607
Db 20431 GGAGCCACACAGCAGGAGGCCACAGCAGCAGGAGGCCACAGCAGCAGGAGGCCACAGCA 20372
Qy 1608 gtgactcgtcagcgtgtcaaaccaatgtcgtcccaacgagatgaactttattc 1667


```
Db 20371 GGAGCCACAGCAGCAGGAGCCACAGAGCGGGAGCCACAGCAGCGGGAGCCCGCAGCAGCG 20312
Qy 1668 gttgatctctagataaacatttaacgtatgcagacattacatcccttttgagccattattt 1727
Db 20311 GGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCCGCAGCAGCG 20252
Qy 1728 aaacaaatcctcaaggatcctgatgcagagagaggtgtacacgtaccatccaaaggagca 1787
Db 20251 GGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCAGGATGAGCAGCAGCA 20192
Qy 1788 cctgtacaagtaccagtgagtagggccgcgcgaagaagtgccacgcgaagaattgatg 1847
Db 20191 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 20132
Qy 1848 caactccaagagacgatttcaattagaagaactgcgaagctccagaggaagagaa 1907
Db 20131 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 20072
Qy 1908 ttagtattagaaggagaaggaaccaaagggaagagccaaagaagagagagcaaca 1967
Db 20071 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCA 20012
Qy 1968 gaagggaagtgcagagaagaattagaggcaactccagagagacgatttcgaattagaa 2027
Db 20011 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19952
Qy 2028 gaaccaacaggagaagtagaagaacccgtagagggcgagaagaactgcagaagagaa 2087
Db 19951 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19892
Qy 2088 gaagtggagaggttacctgcagaagttagaagaagtggagaggtacctgcagaagttagaa 2147
Db 19891 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19832
Qy 2148 gaagtggagaggttaccagaagaagttagaagggttaccgcagaagaatgagaagaagtggaa 2207
Db 19831 GCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCA 19772
Qy 2208 gaggtaccagaagaagtggagaaggttaccagaagaagtggagaaggttaccagaagaagtgg 2267
Db 19771 GCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCA 19712
Qy 2268 gaagggttaccagaagaagtggagaaggttggagaagtagaagaagttagaggttaccagcg 2327
Db 19711 GGAGGAGCAGGAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCA 19652
Qy 2328 gtagtagaagttagaagtaccagcggttagtagaagaagaggtgccagaagaagttagaagaa 2387
Db 19651 GTTAGAGGAGCAGGAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCA 19592
Qy 2388 gaagaagaagagaagaaccagtagaggaagaagtgtattacaattagtaataaccatcg 2447
Db 19591 GGAGTTAGGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCA 19532
Qy 2448 gaagaagatatacaattag 2466
Db 19531 GCAGGAGTTAGAGGAGCAG 19513
```

```
RESULT 8
5231168-1
; Patent No. 5231168
; APPLICANT: DZIEGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO: 1
; LENGTH: 3095
5231168-1
```

```
Query Match 5.2%; Score 175; DB 6; Length 3095;
Best Local Similarity 44.9%; Pred. No. 4.9e-23;
Matches 848; Conservative 0; Mismatches 1020; Indels 21; Gaps 4;

Qy 597 gaaactggcgatttggaaagcaggggaaagaaactggcgatttggaaagcaggggaaagagct 656
Db 16 gaaaaaacgagcatgagcagctgaaatgaagaaagtagcttgaagaaggccatcat 75
Qy 657 ggtgatttggaaagcaggggaaagaaactggcgatttggaaagcaggggaaagaaactggagat 716
Db 76 gaagaattgtacctgaacaaataatgaagaatcgagtgaaagttaaattagttgataat 135
Qy 717 ggggaaactgaagaagagcaactggagatgcggaactgaaatggagcaactgtgatat 776
Db 136 gatgaagggtgttttgaagaagctcatcatgaaaaatttttcatctcgaagtaagtaactct 195
Qy 777 gtgacacagaagatagttcagctgatggagcagaaaaagatacatgttctgtctcaagaa 836
Db 196 gaattaaatgaaaaatgaatttggatctgcacaaagtgtaaactgaacctgtgaacat 255
Qy 837 aatgtacaacctgcgagatgaatgatgcccctcttggaaagtatttggataaagatatata 896
Db 256 gaagaa-----gtgtatctgaagaagcaacctgaaccagctgaaaaatgaagaaagt 309
Qy 897 atttttgatcataattaaagatttcgagccactatttcgacacaaatttggcggtactgct 956
Db 310 agtatagaagaagctcatcaggagaattgtacctgacacaaatgatgaagaatcaggt 369
Qy 957 aaactgtttacgggacaaagtgcctaattgaaacctgtaccattaccagtggcgaagag 1016
Db 370 gaaagtgtattagttgataatgaagaaggtgatttgaagaacctaatcatgaagaattt 429
Qy 1017 ccgcgcgaagtaccagcggaagaattagatgccactccagagagtagacttcgacttagat 1076
Db 430 gaacctgatacaaatgactctgaattgaagtgaataatgatttgaatcagaaaaaagt 489
Qy 1077 gttacagaatctcccgagggaagttagaattagttatgagaagagcgaactggagaagaa 1136
Db 490 gtatctgaaccagctgaacatgtagaattgtatcgaaaaaaagtgtatctgaaccagct 549
Qy 1137 tcaacggaagtgggacccaacggaagagcacaacgcgaagaattagatgccatccagag 1196
Db 550 gaacacgtagaataattgtctgaaaaaagtacatccgaa-----ccagctgaaca 599
Qy 1197 gatggatttcgacttagacgaaactgcagaagagagaacagagaactgtagaggagaa 1256
Db 600 tgcagaaagtgtatctgcacaaagtataacgaaacctccga--aaagaaagatggacca 657
Qy 1257 gaacacagaagaagctgcagaaggagagagtagatcagaagaaactccagaaggagagaag 1316
Db 658 gttccttcaaacatttgaagaaattgaaaaagtgtgttcaacctaaaattgtagac 717
Qy 1317 ttagaggaactccagagagtagatttcgcattagatggaaactacattagagaagaaaccgaa 1376
Db 718 cttcaataattgaacctaattttgtgactcaacccaactccacaagaccagttgaa 777
Qy 1377 gaaactgcagaagagagaaccgttagagggagagaacccgttagaggagagaagaaacc 1436
Db 778 ccatcatttgcataattgaaaagtctcctcgaagaaaataaacaatgcaagtgttgat 837
Qy 1437 gttaggggagaagaagctgcagaaggagagaagagttagaggcaactccagaggagtagac 1496
Db 838 cctgaagtaaaagaaaaaagaaaatgtatctgaagtgttgaagaaaaacaaattcaca 897
Qy 1497 ttccaattagaagaaccatcagaggaagggaagggaagggaaggaagggaagggaagga 1556
Db 898 gaatcagttgaagaataatccagtaaatgaggtgaatttgaaattgaaagtgttcactgaaca 957
Qy 1557 gaaggagaagcgttagtagcagtgccagtagtgccggaaccggttagaagtagtagtacct 1616
Db 958 ttagatttagatcataaaacagttgatccgagaatagtagaagttgaagaataatcctca 1017
```



```
Db 1077 agctccaagtgtgtgtagaaagtgtggtctccaagtgttgaagaagtgttagaagaataatgt 1136
Qy 1247 agagggaagaacacagaagaagctgcagaagaggagaagtatcagaagaataaccacagaagg 1306
Db 1137 tgaagaagtgtagctgaaaaatgttgaagaagtgtagctgaaaaatgttgaagaagtgt 1196
Qy 1307 agaaagaagttagaggcaactccagaggtgatttccattagatggaactacattaga 1366
Db 1197 agctgaaaaatgtgaagaagtgtagctgaaaaatgttgaagaagtgttagctgaaaaatgt 1256
Qy 1367 agaaaccgaagaactgcagaaggggaagaacccgtagagggaagaacccgtagaagg 1426
Db 1257 tgaagaatactgtagctccaactgttgaagaatactgtagctccaactgttgaagaataatgt 1316
Qy 1427 agaaagaacccgtagagggaagaagctgcagaagaagaagaagaagttagaggcaactcc 1486
Db 1317 agctccaagtgttgaagaagtgtggtctccaagtgttgaagaagtgttagaagaataatgt 1376
Qy 1487 agaggtactctccaattagaagaaccatcaggagaaggaagagggaagggaaggaga 1546
Db 1377 tgaagaagtgtagctgaaaaatgttgaagaagtgttagctgaaaaatgttgaagaagtgt 1436
Qy 1547 aggggaaggaggaaggaagcgttagtagcagtgccagtagtggccgaacccgttagaagt 1606
Db 1437 agctgaaaaatgttgaagaagtgttagctgaaaaatgttgaagaagtgttagctgaaaaatgt 1496
Qy 1607 ---agtactcctgctcagcctgtcaaaccaatgctgctccaaaggaaggaaggaacttt 1663
Db 1497 tgaagaagtgtagctgaaaaatgttgaagaagtgttagctgaaaaatgttgaagaataatgt 1556
Qy 1664 attcgttatctcttagaataacgatttaacgttatcgacattacatctcttgaccatt 1723
Db 1557 agctccaactgttgaagaatactgtagctccaactgttgaagaataatgttagctccaagtgt 1616
Qy 1724 atttaacaaactctcaagggaacctgatgcaggaaggctgttaacagtaccatcaaaaga 1783
Db 1617 tgaagaagtgtggtctccaactgttgaagaagtgttagaagaataatgttgaagaagtgt 1676
Qy 1784 agcaactgtacaagtaccagtgccagtagtgccgcgaagaagtgccaacggaagaat 1843
Db 1677 agctgaaaaatgttgaagaagtgtagctgaaaaatgttgaagaagtgttagctgaaaaatgt 1736
Qy 1844 gatcaactccaaggagcatttcgaattagaagaactgcagaagctccagaggaagg 1903
Db 1737 tgaagaagtgtagctgaaaaatgttgaagaatactgtagctccaactgttgaagaataatgt 1796
Qy 1904 agaatttagtattagaagggaagggaacccacggaagaagccaaaggaagggaaggagcc 1963
Db 1797 agctccaactgttgaagaataatgttagctccaactgtttaga-----aagtgtggtcc 1850
Qy 1964 aacagaagaagtgccagaagaagaattagaggcaactccagagcgagcttccgaatt 2023
Db 1851 agtgttgaagaagtgtgaagggaataatgttgaagaagtgttagctgaaaaatgttgaaga 1910
Qy 2024 agaaagaacacagagaagaagttagaagaacccgttagggcggaagaactgcagaagg 2083
Db 1911 agtgttagctgaaaaatgttgaagaagtgttagctgaaaaatgttgaagaataatgttagctcc 1970
Qy 2084 agaaagaagtggaaaggtaccctgcagaagttagaagaagtggaaaggttaccctgcagaagt 2143
Db 1971 aactgttgaagaataatgttagctccaactgttgaagaataatgttagctccaactgtttaga 2030
Qy 2144 agaaagaagtggaaaggtaccagaagaagttagaagaaggtaccgcagaagtagaagaagt 2203
Db 2031 agtgtggtcccaagtgttgaagaagtgttagaagaataatgttgaagaagtgttagctga 2090
Qy 2204 ggaagaggtaccagaagaagtggaaaggtaccagaagaagtggaaaggttaccagaaga 2263
Db 2091 aaatgttgaagaagtgtagctgaaaaatgttgaagaagtgttagctgaaaaatgttgaaga 2150
Qy 2264 agtggaaaggttaccagaagaagtggaaaggtggaaaggttagaagaagttagaggtacc 2323
```

```
Db 2151 aagtgtagctgaaaaatgttgaagaatactgtagctccaactgttgaagaataatgttagctcc 2210
Qy 2324 aqcggttagtagaagttagaagtagtaccagcggttagtagaagaagaaggtgccaagaagaagtaga 2383
Db 2211 aactgttgaagaataatgttagctccaagtgttagaagaagtgtggctcccaaggttgaaga 2270
Qy 2384 aqaagaagaagaaggaaggaaccagtagaggaagaagatgtattacaatttagtaatacc 2443
Db 2271 aagtgtagaagaataatgttgaagaagtgtagctgaaaaatgttgaagaagaagttagctga 2330
Qy 2444 atcgaagaagaatatacaattagacaa 2470
Db 2331 aatgttgaagaagtgttagctgaaaa 2357

RESULT 10
US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1
```

Query Match 5.1%; Score 170.2; DB 4; Length 6152;
Best Local Similarity 44.1%; Pred. No. 4.1e-22;
Matches 911; Conservative 0; Mismatches 1138; Indels 18; Gaps 4;

```
Qy 407 tgagggaagtcacgaacgttagatggaggtgaatcattagggagaggtttccactgaaga 466
Db 550 tgaagaagttaaaggaaaaattcttgactattagaagaaggaatacatattaactgaag 609
Qy 467 tatggaacacagaagatggctcaacagatgatacggaaacacagaagaaggactacctggtga 526
Db 610 tgtagatgataataaaatttagaagaagccgaagataaaaggaataatcttattaag 669
Qy 527 tatggaaggaagaagaagcgtggcgatattggaagcagagggaagaactggcgatttggaaagcag 586
Db 670 taatacagaagaaccaaagtgatcacgaaaaatattatgcacatttataataatttggacaaaa 729
Qy 587 agcaggggaagaactggcgatttggaaagcaggggaagaactggcgatttggaaagcag 646
Db 730 ticagaaaaaacagaagaagtgtatcacgaaaaatgcacagtcagtgatgaacttttaatga 789
Qy 647 ggaagaagctgggtgatttggaaagcaggggaagaactggcgatttggaaagcaggggaaga 706
Db 790 attattaaatagtgtagtgtaattggaagaagtaaaagaaaaatatttggaggaaggtca 849
Qy 707 aactggagatgcggaaactgaagaaggaaggaacactggagatgcggaaactgaaatggagc 766
Db 850 agttaatgcgataatttttaattagtttagtaaaaagtgttcaacaagaacaaacacaaa 909
Qy 767 aactgtgtgttagcacacagaagatagttcagctgtatggagcagagaaaaatcatagtctc 826
Db 910 tgttgaagaataatgtgaagaagtgtagaagaataatgcacgaagaaggttagaagaaaa 969
Qy 827 tgcctgaagaaaaatgtacaacacctcccatagtaatgatgcctcttttggaaaggtattttgga 886
```

Db 970 tgtagaagaaaaatgtagaa-----gaaatgacgcagcgaagtgtgctcctcaagtgtga 1023
QY 887 taaagataataatttgcataataaagatttcgagccactattcgaacaaatttggcc 946
Db 1024 agaaagttagctcaagtgtgatgaaagttagattcaagtattgaagaaaatgtagc 1083
QY 947 gggtaactgctaacaatgttcgaggacaagaattgcgaatgaacacctgtaccattaccagt 1006
Db 1084 tccaactgttgaagaaaatcgtagctcacaagtgt---tgtagaagtgtgctccaagtgt 1140
QY 1007 ggcagaagaccgcgcgaagtaccagcgcgaagaattagatgccactccagagatgactt 1066
Db 1141 tgaagaaagtgtgaagaaataattgtgaagaaagtgtagctgaaatgttgaagaaagtgt 1200
QY 1067 cgcattagattgttacagaatactcccgagggaagttagaattagattagatgaagaggaac 1126
Db 1201 agctgaaaatgttgaagaaagtgtagctgaaaatgttgaagaaagtgttagctgaaaatgt 1260
QY 1127 tgaagaaagaatcaacgcgaagtgggaccacgcgaaggaagggaccacgcgaagaattagatgc 1186
Db 1261 tgaagaaatcgtagctccaactgttgaagaaatcgttagctccaactgttgaagaaatgt 1320
QY 1187 cactccagagatggatttcgcattagcgaataactgcagaagcagaagaaacagaagaacgt 1246
Db 1321 agctcccaagtgtgtagaagaagtgtgctccaagtgttgaagaaagtgtagaagaaatgt 1380
QY 1247 agaggggagaagaacagaagaactgcagaagggagaagtagatcagaagaaactccagaagg 1306
Db 1381 tgaagaaagtgtagctgaaaatgttgaagaaagtgtagctgaaaatgttgaagaaagtgt 1440
QY 1307 agaagaagagttagaggccaactccagaggatgtattcgatttagtggaactacattaga 1366
Db 1441 agctgaaaatgttgaagaaagtgtagctgaaaatgttgaagaaagtgttagctgaaaatgt 1500
QY 1367 agaaaccgaagaactcagaagagagaaacgcgtagaggggagaagaaacccctagagg 1426
Db 1501 tgaagaaatcgtagctccaactgttgaagaaatcgttagctccaactgttgaagaaatgt 1560
QY 1427 agaagaacccgttagagggaagaagaactgcagaagggagaagaagtagtaggcaactcc 1486
Db 1561 agctccaagtgttgaagaagtgtggtcccaagtgttgaagaaagttagaagaaaatgt 1620
QY 1487 agaggatgaactccaattagaagaaccatcaggagaaaggaggaaggggaaggaagaaggaga 1546
Db 1621 tgaagaaagtgtagctgaaaatgttgaagaaagtgtagctgaaaatgttgaagaaagtgt 1680
QY 1547 aggggaagggagaaggaagcgttagtagcagtgccagtagtgccgcgaacccggttagaagt 1606
Db 1681 agctgaaaatgttgaagaaagttagctgaaaatgttgaagaaagtgttagctgaaaatgt 1740
QY 1607 ---agtgactcctgctcagcctgtcacaaccatgttgctgccacgcgcagatgaaacttt 1663
Db 1741 tgaagaaagtgtagctgaaaatgttgaagaaagttagctgaaaatgttgaagaaatcgt 1800
QY 1664 attcgttgatcttagataacgatttaacgtatgcagacattacatccttttagccatt 1723
Db 1801 agctccaactcgttgaagaaatcgtagctccaactgttgaagaaatgttagctccaagtgt 1860
QY 1724 atttaacaaatcctcaagatcctgatgcaggagagcgtgtaacagtagccatccaaggga 1783
Db 1861 tgaagaaagtgtggtcctccaagtgttgaagaaagttagaagaaaatgttgaagaaagtgt 1920
QY 1784 agaacctgtcacagtagaccagtggcagtagggccgcgcgaagagtgccacgcgaagaatt 1843
Db 1921 agctgaaaatgttgaagaaagttagctgaaaatgttgaagaaagttagctgaaaatgt 1980
QY 1844 gatgcaactccaagagacgatttcgaattagaagaactgcagaagctccagaggaagg 1903
Db 1981 tgaagaaagtgtgagctgaaaatgttgaagaaatcgttagctccaactgttgaagaaatcgt 2040
QY 1904 agaattagattagaagaggaaggaacaaacgcggaagaaagagccagagaaagagagcc 1963
Db 2041 agctccaactgttgaagaaatgttagctccaagtgtttaga-----aagtgtggctcc 2094

QY 1964 acagaaggagaaagtgccagaagaagaatttagaggcgaactccagagacgatttcgaatt 2023
Db 2095 aagtgttgaagaaagttagaagaaaatgttgaagaaagttagctgaaaatgttgaaga 2154
QY 2024 agaagaaccaacaggagaagaagaatagaagaaaccgttagagggcgaagaaactgcagaagg 2083
Db 2155 aagttagctgaaaatgttgaagaaagttagctgaaaatgttgaagaaatcgttagctcc 2214
QY 2084 agaagaagtgaagaggtacctgcagaagttagaagaagtgaagaggtacctgcagaagt 2143
Db 2215 aactgttgaagaaatcgttagctccaactgttgaagaaatgttagctccaagtgtttaga 2274
QY 2144 agaagaagtgaagaggtaccagaagaagaatagaagaggttccccgcgaagaatagaagaagt 2203
Db 2275 aagtgtggctcccaagtgttgaagaaagttagaagaaaatgttgaagaaagttagctga 2334
QY 2204 ggaaggggtaccagaagaagtggagaaggttaccagaagaagtggagaaggtaccagaaga 2263
Db 2335 aatgttgaagaaagttagctgaaaatgttgaagaaagttagctgaaaatgttgaaga 2394
QY 2264 agtggaaagaggttaccagaagaagtggagaagtggagaagtagaagaagttagaggtacc 2323
Db 2395 aagttagctgaaaatgttgaagaaatcgttagctccaactgttgaagaaatcgttagctcc 2454
QY 2324 agcggttagtaagaagtaccagcggtagtagaagaagaaggtgccagaagaagttaga 2383
Db 2455 aactgttgaagaaatgttagctccaagtgttagaaaagtgtggtcccaagtgttgaaga 2514
QY 2384 agaagaagaagaagaagaaccagtagagggaagaagatgtattacaattagtaatacc 2443
Db 2515 aagttagaagaaaatgttgaagaaagttagctgaaaatgttgaagaaagttagctga 2574
QY 2444 atcggaagaagatatacaattagacaa 2470
Db 2575 aatgttgaagaaagttagctgaaa 2601

RESULT 11
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 6114111
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hiang, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/050,863
; FILING DATE: 30-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:

Qy	2219	agaagtggaaaggtaccagaagaagtggaaaggtaccagaagaagtggaaaggtacc	2278
Db	1109	AGGAGGGCGCAGGACAGGAGGGCGCAGGAGGGCGCAGGAGGGCGCAGGAGGGCGAGG	1168
Qy	2279	agaagaagtgaagaagtgaagaadtgaagaagttagaaggtaccacgcggtagtgaagt	2338
Db	1169	AGCAGGAGGGCGCAGGAGGGCGCAGGAGCGAGGAGGGCGCAGGAGGGCGCAGGAGGGCGAGG	1228
Qy	2339	agaagtaccagcggtagtgaagaagagtgccagaagaagttagaagaagaagaagaaga	2398
Db	1229	AGGGCGAGGAGGGCGCAGGACGAGAGGGCGCAGGAGGGCGCAGGAGGGCGCAGGAGGGCGAGG	1288
Qy	2399	ggagaaccagttagagaagaagatgtattacaattagtaataccatcgcgaagaag	2454
Db	1289	GGCAGGAGCAGGAGGGCGCAGGAGGGCGCAGGAGGGCGCAGGAGGGCGCAGGAGGGCGAGG	1344

RESULT 13
 US-09-130-114-1/c
 ; Sequence 1, Application US/09130114
 ; Patent No. 5976807
 ; GENERAL INFORMATION:
 ; APPLICANT: Horlick, Robert A.
 ; APPLICANT: Damaj, Bassam B.
 ; APPLICANT: Robbins, Alan K.
 ; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
 ; TITLE OF INVENTION: From Multiple Transfected Episomes
 ; FILE REFERENCE: 0867/10903U51
 ; CURRENT APPLICATION NUMBER: US/09/130,114
 ; CURRENT FILING DATE: 1998-08-06
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 5452
 ; TYPE: DNA
 ; ORGANISM: VEBNA
 US-09-130-114-1

Query Match	5.0%	Score 167.2;	DB 2;	Length 5452;
Best Local Similarity	52.1%;	Pred. No. 1.4e-21;		
Matches 373;	Conservative 0;	Mismatches 343;	Indels 0;	Gaps 0;
Qy 1739	caaggatcctgatgcagagagggctgttaacagtaccatcaaaaggaagcacctgttacaagt	1798		
Db 2176	CAMAGGGACCCACGGTGTGAACAGGNCAGGACAGGACGGGAGGGGACGAGGACGAGGAGG	2117		
Qy 1799	accagtgtagtagggcccgcgcaagaagtgccaaacggaagaattgtatgtaacctccaaga	1858		
Db 2116	GGCAGGACGAGGAGGGGGCAGGACGACGAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG	2057		
Qy 1859	ggaagatttcgaattagaaggaactgcagaagctccagaagctccagaaggaagagataattagttataga	1918		
Db 2056	AGGGGCAGGACGAGGACGAGGGGGCAGGAGCAGGACGAGGAGGGGCAGGAGGGGCAGGAGGGGC	1997		
Qy 1919	aggagaagtgaaaccaacgcggaagaagccaaagagaagagagccaaacagaagagagaagt	1978		
Db 1996	AGGACGAGGAGGGGCAGGACGACGAGGAGGAGGGGCAGGAGGGGCAGGACGAGGAGGAGGAGG	1937		
Qy 1979	gccagaagaagaattagaggcaactccagagagagcagattctcgaattagaagaacccaacagg	2038		
Db 1936	GGCAGGAGGGGCAGGAGGGGGCAGGACGACGAGGAGGAGGGGCAGGACGAGGAGGGGCAGG	1877		
Qy 2039	agaagaagttagaagaacccgttagaggccgaagaactgcagaaggaagagaagtgggaaga	2098		
Db 1876	AGGGGCAGGACGAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGACGAGGAGGAGGGGC	1817		
Qy 2099	ggtacctgcagaagttagaagaagtggaaagaggttacctgcagaagttagaagaagtgggaaga	2158		
Db 1816	AGGACGAGGGGCAGGAGGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGAGG	1757		
Qy 2159	ggtaccagaagaagttagaagaaggttaccgcgcagaagtagaagaagtggaaagaggttaccaga	2218		

Db	1756	AGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGC	1697
Qy	2219	agaagtgaagaggtaccagaagaagtgaagaggtaccagaagaagtgaagaggtacc	2278
Db	1696	AGGAGGGGCAGGACAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG	1637
Qy	2279	agaagaagtgaagaagtgaagaagtgaagaagtgaagaagtgaagaagtgaagaagt	2338
Db	1636	AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGG	1577
Qy	2339	agaagtaccagcgttagaagaagaaggtccagaagaagtgaagaagaagaagaaga	2398
Db	1576	AGGGCAGGAGGGGCAGGACAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGG	1517
Qy	2399	ggaagaaccagtgaagaagaagtattacaattagtaataaccatcgcgaagaag	2454
Db	1516	GCCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGG	1461

```

RESULT 14
US-08-910-647-1
: Sequence 1, Application us/08910647
: Patent No. 6251433
: GENERAL INFORMATION:
: APPLICANT: Zuckermann et al.
: TITLE OF INVENTION: Compositions and Methods for
: TITLE OF INVENTION: Polynucleotide Delivery
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 94608-2916
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/910,647
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Fujita, Sharon M.
: REGISTRATION NUMBER: 38,459
: REFERENCE/DOCKET NUMBER: 1218.002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 923-2706
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9600 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-910-647-1

```

	Query Match	5.0%	Score 167.2;	DB 4;	Length 9600;
	Best Local Similarity	52.1%;	pred. No. 1.6e-21;		
	Matches 373;	Conservative 0;	Mismatches 343;	Indels 0;	Gaps 0;
Qy	1739	caaggatcctgatgcaggagagcgctgttaacagtaccatacaaaggaagcacctgtacaagt	1798		
Dd	675	CAAGGGACCCACGCTGTGAACAGGACGAGGACGAGCGGAGGGGCAGGACGAGGAGG	734		
Qy	1799	accagtgscagttagggcccgcgcagaagtgccaacgaaaaattgtatgcactccaaga	1858		
Dd	735	GGCAGGACGAGGAGGGGGCAGGACGACGAGGAGGAGGGGCAGGAGGGGCAGG	794		


```

; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;
; NUMBER OF SEQUENCES: 21
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
;
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194.087
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
;
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
;
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-194-087-15

```

Query Match	5.0%	Score 167.2;	DB 2;	Length 10596;
Best Local Similarity	52.1%	Pred. No. 1.6e-21;		
Matches 373;	Conservative	0;	Mismatches 343;	Indels 0;
Gaps	0;			
QY	1739	caagatccctgatgcagagagcgctgtaacagtaccatcaaaaggaagcacctgtacaagt	1798	
Db	2209	CAAAAGGACCCACCGTGGACAGGACGAGGACGAGGCGGAGCGGCACGAGCAGGAGG	2268	
QY	1799	accagtgccagtagggcccgcgcaagaagtgcacaacggaagaattgatgcacctccaaga	1858	
Db	2269	GGCAGGACGAGGAGGCGGACGACAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG	2328	
QY	1859	ggacgatttcgaattagaaggaactgcagaagctccagaggaagagagaattagttattaga	1918	
Db	2329	AGGGGCAGGACGAGGAGGCGGCAGGACGAGGAGGAGGGGCAGGAGGGGCAGGAGGGGC	2388	
QY	1919	aggagaaggaagacccaacggaagaagccaagaagaagagagcccaacagagaaggaagt	1978	
Db	2389	AGGAGCAGGAGGAGGGGCAGGACGAGGAGGAGGGGCAGGAGGGGCAGGAGGAGGAGG	2448	
QY	1979	gcagaagaagaattagagccaactccagagagacgatttcgaattagaagaaccaacagg	2038	
Db	2449	GGCAGGAGGGGCAGGAGGGGCAGGACGAGGAGGAGGGGCAGGAGGAGGAGGGGCAGG	2508	
QY	2039	agaagaagtagaagaaccgttagaggcgcaagaactgcagaagaagagaagaagtggaaag	2098	
Db	2509	AGGGGCAGGACGAGGAGGCGGCAGGAGGGGCAGGAGGGGCAGGACGAGGAGGAGGGGC	2568	
QY	2099	ggtacctgcagaagttagaagaagtgggaagaggttacctgcagaagtagaagaagtggaaaga	2158	
Db	2569	AGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGAGGAGG	2628	
QY	2159	ggtaccagaagaagtagaagaaggttaccocgcagaagtagaagaagtggaagaaggtaccaga	2218	
Db	2629	AGGGGCAGGAGGGGCAGGAGGGGCAGGACGAGGAGGAGGGGCAGGAGGAGGAGGGGCAGG	2688	

[illegible]

	Query Match	5.08;	Score 167.2;	DB 5;	Length 10596;
	Best Local Similarity	52.1%;	Pred. No. 1.6e-21;		
	Matches 373;	Conservative	0; Mismatches 343;	Indels	0; Gaps
Qy 1739	caaggatcctgatcgadgagaggggtgtaacagtaccatcaagaagcacctgtacaagt				
Db 2209	CAAAAGGACCCACCGTGGTAACAGGAGCAGGAGCAGGAGCGGGGAGGACGAGG				
Qy 1799	acagtgtagcatagggccgcqcaagaagtgtccaacggaagaattgtatgcactccaaga				

Db 3787 GAGAGGCGCGATTGGCTGAAGCGCGAGGAGGAGGAGCGGAGGAGCGG 3846
Qy 1888 aagctccagaggaagagattatattagaaagagagagagaaacaaagaaagc 1947
Db 3847 AAGCAGCGTGAAGCCGACCGAGGAGGAGGAGGCGGCGGCTTGAAGCTCGAG 3906
Qy 1948 caagagaagagagcaacagaagagaagtgccagaagaagattagaggcaactccag 2007
Db 3907 GAGGAGGAGGCGCGAGCAGCAGGAGGCGCGGAGCAGCACTAAGCGGGAGCAAGAG 3966
Qy 2008 aggacgatttogaattagaagaacaaagaggaagaagttagaagaacccgttagaggcg 2067
Db 3967 GAGAGGCGCGAGCAGCGCTGAAGCGCGAGGAGGAGGAGGCGGCTCCAGCAGCGGTTG 4026
Qy 2068 aagaaactgc-----aagagaagaagaagtgaagaggtaccctgcagaag 2112
Db 4027 AGGAGCAGCAACAACTAGACGCGAGCAGGAGGAGGCTCGAGCAGCTGCTGAAGCGC 4086
Qy 2113 tagaagaagtggaaagaggtacctgcagaagttagaagaagtggaaaggtaccagaagaag 2172
Db 4087 GAGGAGGAGAGAGCTCGAGCAGGAGGCGGAGCAGCGGCTGAAGCGGAGCAGGAG 4146
Qy 2173 tagaagaaggtaccgcgagaagtagaagaagtggaaaggtaccagaagaagtgaaggg 2232
Db 4147 GAGAGGCGCGATCAGCTGCTGAAGCGCGAGGAGGAGGAGGCGGCTGAAGCGC 4206
Qy 2233 taccagaagaagtgaagaggtaccagaagaagtggaaaggtaccagaagaagtgaag 2292
Db 4207 GAGCAGGAAGAGAGCTCGAGCAGCGACTGAAGCGCGAGGAGGCTGAGAGCTGCGAGCAG 4266
Qy 2293 aagtggaaagtagtaagaagtagaggtaccagcggtagtagaagtagaagtagaccagcg 2352
Db 4267 GAGGAGGCGCGAGCAGCGGCTGAAGCGCGAGGAGGCGGAGGAGGCGGCGCACGAG 4326
Qy 2353 tagtagaagaagaggtccagagaagaagttagaagaagaagaagagagaagaaccagtag 2412
Db 4327 CTGCTGAAGAGCGGAGGAGGAGGAGGAGGCGGCGCACGAGCAACTGAGGCGCGAGCAGCAG 4386

RESULT 22

US-08-800-644-93
; Sequence 93, Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:

; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9551 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1507..1644
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1645..2511
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2512..8070
; US-08-800-644-93

Query Match 4.4%; Score 147.2; DB 2; Length 9551;
Best Local Similarity 47.2%; Pred. No. 5.9e-18;
Matches 586; Conservative 0; Mismatches 613; Indels 21; Gaps 3;

Qy 1231 gaacacagaagaacagtagaggagagaagaacagaagaagctgcagaaggagagaagtacag 1290
Db 3190 GAGGAAGAGCAGCAGCTACGGAAGCTGGAGCGCGCAAGAGCTGAGGAGGGAGCGCAGGAG 3249
Qy 1291 aagaaactccagaaagagaagaagtagtagagcaactccagaggatgatttcgattag 1350
Db 3250 GAAGAGCAGCAGCAGCAAGAGGCTGAGGCGCGAGCAGCACTAAGGGCGCAAGCAGGAGGAG 3309
Qy 1351 atggaactacattagaaagaaacccagaagaaactgcagaagagaagaagaaacgcttagaggag 1410
Db 3310 GAGAGGCGCGCAGCAGCAGGAGGAGGCGCGCAGCAGCAGGAGGCGCGCAGCAGCAGGAG 3369
Qy 1411 aagaaacccgttagaggagagaacccgttagaggagagaagaagctgcagaaggagagaag 1470
Db 3370 GAGAGGCGCGCAGCAGCAGCTGAGGCGCGAGGAGGAGGAGGCGCGCAGCAGCAGCTGAGG 3429
Qy 1471 agttagaggcaactccagagagtagtccattccattagaagaaccatcagagaagagaag 1530
Db 3430 CGCGAGCAGGAGGAGGAGGCGCGCAGCAGCTGAGGCGCGCAGGAGGAGGAGGAGGAGG 3489
Qy 1531 ggaagagagaaggaaggggagagaagaagcgttagtagcagtgccagtagtg 1590
Db 3490 CGCGCAGCAGCAGCTGAGGCGCGAGCAGGAGGAGGAGGCGCGCAGCAGCAGCTGAGGCGC 3549
Qy 1591 ccgaacccggttagaagtagtgactctctgcagcctgtcaaaccaatggtcgctccacag 1650
Db 3550 GAGCAGCAGCTGAGGCGCGCAGCAGCAGCTGAGGCGCGCAGCAGCAGCTGAGGCGCGCAGCAG 3609
Qy 1651 cagatgaactttatttcgttgatatcttagataacatttaacgtatgcagacattacat 1710
Db 3610 CAGCTGAGGCGCGCAGCAGCAGCT---GAGGCGCGCAGCAGCAGCTGAGGCGCGCAGCAGCAG 3666
Qy 1711 cctttgagccattattttaacaaatcctcaagatcctgatgcagaggagaggtgttaacag 1770
Db 3667 CTGAGGCGCGCAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGCAGCAGGAGGAGGAG 3726
Qy 1771 tacatcaaaaggagaacacctgtacaagtaccagtggcagtagggcccgccagagaagtgc 1830
Db 3727 AGGCACGAGCAGAGCAGCAGGAGGAGGCGCGGCTGAAGCGCGCAGCAGGAGGAG 3786
Qy 1831 caacggaagaattgatgcaactccaagagagacgatttcga---attagaaggaaactgcag 1887


```
QY 1940 agaagagccaagagaagagagccacagaagagagagtagtccagagaagaattagggc 1999
Db 168 agatgccacagaaaaagggagagaaaaagaaagcagtgccagcagataaaaaatga 227
QY 2000 aactccagagacgatttcgaattagaagaacccaacagagagaagtagaagaaccgct 2059
Db 228 agaagaagatcagaagaagatgaagaagatcaaaacgaagagaaggggaagctggaaa 287
QY 2060 agagggcgaaacactgcagaagagaagaagtggaaaggtaccctgcagaagtagaaga 2119
Db 288 agaagacaagatgaaaaaggggaagagatggaaagagga taaaaatggaaatagaa 347
QY 2120 agtggaaagagtgacctgcagaagtagaagaagtggaaaggtaccagagaagaatagaaga 2179
Db 348 aggaagagatgcaaaagagaagaagaagtggaaaaaagtgaaagacggaaaggaatgg 407
QY 2180 ggtaccgcagaagtagaagaagtggaaaggtaccagaagaagtggaagaggtaccaga 2239
Db 408 agaagatggaaaagagaagaagatgaaaaaggggaagagacagaaaaagacagg 467
QY 2240 agaagtggaagaggtaccagaagaagtggaaaggtaccagaagaagtggaaagtgga 2299
Db 468 agttgaaaagagaatgaagatggaaaagagaagggagataaaaaagagggggaagatgt 527
QY 2300 agaagtagaagaagtagaggtaccagcggtagtagaagtagaagtagtaccagcggtagtaga 2359
Db 528 aaaagtcaaaagagatgaaaagagagagaagatggaaaagagatgaagtggaatga 587
QY 2360 agaagaagtcagagaagaagtagaagaagaagaagaagaagaagacacagtagagaaga 2419
Db 588 ggaagaagctggaaaagagaagaagatttaaaagagaagagagaaggaaggaaga 647
QY 2420 agatgtattacaa 2432
Db 648 tgagatcaaaaga 660

RESULT 25
US-09-461-697-189
; Sequence 189, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 189
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-189
```

```
Query Match 4.3%; Score 145; DB 4; Length 717;
Best Local Similarity 53.9%; Pred. No. 8.2e-18;
Matches 298; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

QY 1880 aactgcagaagctccagagaagagaattagtagaaggaaggaaggaacacagga 1939
Db 126 aattacagagggcaccagcttctgaaaaagaattgtggaagtaaaagaagaataattga 185
QY 1940 agaagaccaagagaagagagccacagaagagagaagtagtccagagaagaattagaggc 1999
```

```
Db 186 agatgccacagaaaaagggagagaaaaagaaagcagtgccagcagataaaaaatga 245
QY 2000 aactccagagacgatttcgaattagaagaacccaacagagagaagtagaagaaccgct 2059
Db 246 agaagaagatcagaagaagatgaagaagatcaaaacgaagagaaggggaagctggaaa 305
QY 2060 agagggcgaaacactgcagaagagaagaagtggaaaggtaccctgcagaagtagaaga 2119
Db 306 agaagacaagatgaaaaaggggaagaagatggaaaagagggataaaaaatggaaatgagaa 365
QY 2120 agtggaaagagtgacctgcagaagtagaagaagtggaaaggtaccagagaagaatagaaga 2179
Db 366 aggaagagatgcaaaagagaagaagaagtggaaaaaagtgaaagacggaaaggaatgg 425
QY 2180 ggtaccgcagaagtagaagaagtggaaaggtaccagaagaagtggaagaggtaccaga 2239
Db 426 agaagatggaaaagagaagaagatgaaaaaggggaagatgaaaaaggggaagacagaaaaacagg 485
QY 2240 agaagtggaagaggtaccagaagaagtggaaaggtaccagaagaagtggaaagtgga 2299
Db 486 agttgaaaagagaatgaagatggaaaagagaagggagataaaaaagagggggaagatgt 545
QY 2300 agaagtagaagaagtagaggtaccagcggtagtagaagtagaagtagtaccagcggtagtaga 2359
Db 546 aaaagtcaaaagagatgaaaagagagagaagatggaaaagagatgaagtggaatga 605
QY 2360 agaagaaggtgccagaagaagtagaagaagaagaagaagaaggaacacagtagagaaga 2419
Db 606 ggaagaagctggaaaagagaagaagatttaaaagagaaggggaaggaaggaaga 665
QY 2420 agatgtattacaa 2432
Db 666 tgagatcaaaaga 678
```

```
RESULT 26
US-09-461-697-187
; Sequence 187, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187
; LENGTH: 774
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-187
```

```
Query Match 4.3%; Score 145; DB 4; Length 774;
Best Local Similarity 53.9%; Pred. No. 8.3e-18;
Matches 298; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

QY 1880 aactgcagaagctccagagaagagaattagtagaaggaaggaaggaacacagga 1939
Db 183 aattacagagggcaccagcttctgaaaaagaattgtggaagtaaaagaagaataattga 242
QY 1940 agaagaccaagagaagagagccacagaagagagaagtagtccagagaagaattagaggc 1999
```


Db 243 agatgccacagaaaggaggagaaagaaagcagtgccagcagaagtataaaatga 302
QY 2000 aactccagaggacgatttcgaaattagaagaaccacacagagagaagtagaagaacccgt 2059
Db 303 agaagaagatcagaagaagatgaagaatcaaacgaagaaagggaagctggaaa 362
QY 2060 agagggccgaagaactgcagaaagagaagaagtggaaagggtagctgcagaaagtagaaga 2119
Db 363 agaagacaagatgaaaaagggaagaaagatggaaaagaggaataaaatggaaatgaga 422
QY 2120 agtggaaaggtaccctgcagaagtacaagaagtggaaaggtaccagaagaagtagaaga 2179
Db 423 agagaagaatgcgaagagaaagaagatggaataaagggtgaagacggaaggaatgg 482
QY 2180 ggtaccgcgagaaagtgaagaagtggaagagggtaccagaagaagtggaagaggtaccaga 2239
Db 483 agaagatggaagaagaaaggagaagatgaaaagaggaagacagacagaaagaacagg 542
QY 2240 agaagtggaaagaggtaccagaagaagtggaaaggggtaccagaagaagtggaagaagtgg 2299
Db 543 agtggaaaagagaatgaagatggaaaagagaaggagataaaaagaggggaaagatgt 602
QY 2300 agaagtagaagaagtagaggtaccagcggtagtagaagtagaagtagaccagcggtagtaga 2359
Db 603 aaagtcacaagaagatgaaaagagagagaagatggaagaagaagatgaaggtggaatga 662
QY 2360 agaagaggtgccagaagaagtagaagaagaagaagagaagagaagacacagtagagaaga 2419
Db 663 ggaagaagctggaagaagaagaagaattttaaaagagaggaaggaaggaaggaaga 722
QY 2420 agatgtattacaa 2432
Db 723 tgagatcaaaaga 735

RESULT 27
US-09-461-697-185
; Sequence 185, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 185
; LENGTH: 819
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-185

Query Match 4.3%; Score 145; DB 4; Length 819;
Best Local Similarity 53.9%; Pred. No. 8.4e-18;
Matches 298; Conservative 0; Mismatches 255; Indels 0; Gaps 0;
QY 1880 aactcgagaagctccagagaagaagaattagtagtagaaggagaaggaacacagga 1939
Db 228 aattacagagccagcgtcttgaaaagaagaattgtggaagtaaaagaagaataattga 287
QY 1940 agaagaccaagaagagagagccacagagaagagaagtagtccagagaagaagaattagaggc 1999
Db 288 agatgccacagaaaggaggagaaagaagaagcagtcagtcgagcagaagtaaaaaatga 347

QY 2000 aactccagaggacgatttcgaaattagaagaaccacacagagagaagtagaagaacccgt 2059
Db 348 agaagaagatcagaagaagatgaagaatcaaacgaagagaagaagggaagctggaaa 407
QY 2060 agagggccgaagaactgcagaaagagaagaagtggaaagggtagctgcagaaagtagaaga 2119
Db 408 agaagacaagatgaaaaagggaagaaagatggaaaagaggaataaaatggaaatgaga 467
QY 2120 agtggaaaggtaccctgcagaadtacaagaagtggaaaggtaccagaagaagtagaaga 2179
Db 468 agagaagaatgcgaagagaaagaagatggaaaaggggtgaagacggaaggaatgg 527
QY 2180 ggtaccgcgagaaagtgaagaagtggaagagggtaccagaagaagtggaagaggtaccaga 2239
Db 528 agaagatggaagaagagaagagagaagatgaaaagaggaagacagacagaaagaacagg 587
QY 2240 agaagtggaaagaggtaccagaagaagtggaaaggggtaccagaagaagtggaagaagtgg 2299
Db 588 agtggaaaagagaatgaagatggaaaagagaaggagataaaaagaggggaaagatgt 647
QY 2300 agaagtagaagaagtagaggtaccagcggtagtagaagtagaagtagaccagcggtagtaga 2359
Db 648 aaagtcacaagaagatgaaaagagagagaagatggaagaagaagatgaaggtggaatga 707
QY 2360 agaagaggtgccagaagaagtagaagaagaagaagaagagaaggaagacacagtagagaaga 2419
Db 708 ggaagaagctggaagaagaagaagaattttaaaagagaggaaggaaggaaggaagaaga 767
QY 2420 agatgtattacaa 2432
Db 768 tgagatcaaaaga 780

RESULT 28
US-09-461-697-184
; Sequence 184, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 184
; LENGTH: 1669
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-184

Query Match 4.3%; Score 145; DB 4; Length 1669;
Best Local Similarity 53.9%; Pred. No. 9.9e-18;
Matches 298; Conservative 0; Mismatches 255; Indels 0; Gaps 0;
QY 1880 aactcgagaagctccagagaagaagaattagtagtagaaggagaaggaacacacagga 1939
Db 307 aattacagagccagcgtcttgaaaagaagaattgtggaagtaaaagaagaataattga 366
QY 1940 agaagaccaagaagagagagccacagagaagagaagtagtccagagaagaagaattagaggc 1999
Db 367 agatgccacagaaaggaggagaaagaagaagcagtcagtcgagcagaagtaaaaaatga 426

Search completed: June 14, 2002, 15:29:39
Job time: 19070 sec

Result No.	Score	Query Match	Length	DB	ID	Description
1	3337	100.0	3337	17	AAI34620	P. vivax ESP-1 blo
2	3337	100.0	3337	20	AAI51174	DNA encoding a sec
3	3337	100.0	3337	22	AAH76457	Plasmodium vivax E
4	342.8	10.3	3399	17	AAI05868	Chicken leucocytoz
5	268	8.0	1686	16	AAQ87587	DNA encoding Leuco
6	228	6.8	3579	21	AAAF00999	Plasmodium falcipa
7	220.2	6.6	2805	22	AAQ071130	Canine retinitis p
8	219	6.6	2019	24	ABA30848	Bacillus anthracis
9	217.4	6.5	2849	22	AAQ21684	Human retinitis pi


```
QY 1741 aggatcctgatgcaggagaggtgttaacagttaccatacaaggaagcacctgtacaagtac 1800
|||||
Db 1741 aggatcctgatgcaggagaggtgttaacagttaccatacaaggaagcacctgtacaagtac 1800

QY 1801 cagtggcagtagggccgcgcaagaagtcccaacggaagaattgatgcaactcccaagagg 1860
|||||
Db 1801 cagtggcagtagggccgcgcaagaagtcccaacggaagaattgatgcaactcccaagagg 1860

QY 1861 acgatttcgaattagaaagaactgcagaagctccagaggagaagaattagttattagaag 1920
|||||
Db 1861 acgatttcgaattagaaagaactgcagaagctccagaggagaagaattagttattagaag 1920

QY 1921 ggaagagaaaccaacggaagagcgaagagagagagagagcacaacagagagagagtgctc 1980
|||||
Db 1921 ggaagagaaaccaacggaagagcgaagagagagagagagcacaacagagagagtgctc 1980

QY 1981 cagaagaagaattagaggcaactccagagagcgaatttcgaattagagaaccaacagaggag 2040
|||||
Db 1981 cagaagaagaattagaggcaactccagagagcgaatttcgaattagagaaccaacagaggag 2040

QY 2041 aagaagtagaacaacccgtagaggcggaagaactgcagaagagagaagaagtggaaagg 2100
|||||
Db 2041 aagaagtagaacaacccgtagaggcggaagaactgcagaagagagaagaagtggaaagg 2100

QY 2101 tacctgcagaagttagaagaagtggaaagggttacctgcagaagttagaagaagtggaaagg 2160
|||||
Db 2101 tacctgcagaagttagaagaagtggaaagggttacctgcagaagttagaagaagtggaaagg 2160

QY 2161 taccgaagaagaatagaagaagggtaccgcgagaagtagaagaagtggaaagggttaccgaagg 2220
|||||
Db 2161 taccgaagaagaatagaagaagggtaccgcgagaagtagaagaagtggaaagggttaccgaagg 2220

QY 2221 aagtgaagcagtagcagagaagaagtggaaagggtaccgaagaagtggaaagggtaccag 2280
|||||
Db 2221 aagtgaagcagtagcagagaagaagtggaaagggtaccgaagaagtggaaagggtaccag 2280

QY 2281 aagaagtggaagaagtggaaagtagaagaagtagaggtaccagcggtagtagaagtag 2340
|||||
Db 2281 aagaagtggaagaagtggaaagtagaagaagtagaggtaccagcggtagtagaagtag 2340

QY 2341 aagtaccagcggtagtagaagaagaaggtgccagagaagaatagaagaagaagaagaagg 2400
|||||
Db 2341 aagtaccagcggtagtagaagaagaaggtgccagagaagaatagaagaagaagaagaagg 2400

QY 2401 aagaaccagtagagaagaagaatgtattacaattagtaaccatcggaagaagatatatc 2460
|||||
Db 2401 aagaaccagtagagaagaagaatgtattacaattagtaaccatcggaagaagatatatc 2460

QY 2461 aattagacaaaccaagaagaacgaattaggctctggaatttttatctatcatcgacatgc 2520
|||||
Db 2461 aattagacaaaccaagaagaacgaattaggctctggaatttttatctatcatcgacatgc 2520

QY 2521 actacaagaacgttccaagaagaattttatggaagaagaagaagaactgcagtgatccat 2580
|||||
Db 2521 actacaagaacgttccaagaagaattttatggaagaagaagaagaactgcagtgatccat 2580

QY 2581 tgaaccagaagaattttgcaaggaagaattccaaactcagaagaatggtccacattcattc 2640
|||||
Db 2581 tgaaccagaagaattttgcaaggaagaattccaaactcagaagaatggtccacattcattc 2640

QY 2641 aaggcctagaagggcactgggaacgattagaagtgagcttaataaaggctagagaagaat 2700
|||||
Db 2641 aaggcctagaagggcactgggaacgattagaagtgagcttaataaaggctagagaagaat 2700

QY 2701 ggaatgaacaagaataaagaatggcgtggctggcttcgttaattgaaataaattggt 2760
|||||
Db 2701 ggaatgaacaagaataaagaatggcgtggctggcttcgttaattgaaataaattggt 2760

QY 2761 cagaatatagtcataattcaacaaaggaagaccacagctggtttggaagaacagagat 2820
|||||
Db 2761 cagaatatagtcataattcaacaaaggaagaccacagctggtttggaagaacagagat 2820
```

```
QY 2821 ggagcgcagagaaatgtaaaaaatggttttaaaagcagaagtcacaaatgccaaattgattcac 2880
|||||
Db 2821 ggagcgcagagaaatgtaaaaaatggttttaaaagcagaagtcacaaatgccaaattgattcac 2880

QY 2881 acttgaaaaaatggatgaaacgcacactcattccattatttataaaattcttctgtgaagata 2940
|||||
Db 2881 acttgaaaaaatggatgaaacgcacactcattcccaattatttataaaattcttctgtgaagata 2940

QY 2941 tgtcacaaatttgaaaaaaagaacaaacaaagaatgggttaataatgaaactggaaaaaaacg 3000
|||||
Db 2941 tgtcacaaatttgaaaaaaagaacaaacaaagaatgggttaataatgaaactggaaaaaaacg 3000

QY 3001 aacgggggttatgggttctgaatcatttgaagttatgaccacatcaaaattatttaaatgtgg 3060
|||||
Db 3001 aacgggggttatgggttctgaatcatttgaagttatgaccacatcaaaattatttaaatgtgg 3060

QY 3061 ctaagagtcgagaatgggtaccgtgccaatctctaataatagagaagaaggaactca 3120
|||||
Db 3061 ctaagagtcgagaatgggtaccgtgccaatctctaataatagagaagaaggaactca 3120

QY 3121 tgaatgggttctcctaaaaaagaaacgaatattttagacaaagaatggaaaaatggact 3180
|||||
Db 3121 tgaatgggttctcctaaaaaagaaacgaatattttagacaaagaatggaaaaatggact 3180

QY 3181 cattggaaaaaagttaaaatttttggttcaattcgaatgtgtacaacattttctggaata 3240
|||||
Db 3181 cattggaaaaaagttaaaatttttggttcaattcgaatgtgtacaacattttctggaata 3240

QY 3241 cgcttaacaaaggaagaatggaaatcaatttggtaatttgaataaaagtttgaattataga 3300
|||||
Db 3241 cgcttaacaaaggaagaatggaaatcaatttggtaatttgaataaaagtttgaattataga 3300

QY 3301 aaaaagacagattctctctataaaataaaataattc 3337
|||||
Db 3301 aaaaagacagattctctctataaaataaaataattc 3337

RESULT 2
AAK15174
ID AAK15174 standard; DNA; 3337 BP.
AC AAK15174;
XX
DT 28-APR-1999 (first entry)
XX
DE DNA encoding a secreted blood-stage protein called PvESP-1.
KW Erythrocyte secreted protein-1; PvESP-1; malarial antigen;
XX blood-stage protein; malaria; monoclonal antibody ID11G10; ds.
OS Plasmodium vivax.
XX
FH Key
CDS 1..3197
FT /*tag= a
FT /*note= "contains 1 intron"
FT exon 1..91
FT /*tag= b
FT /*number= 1
FT 92..230
FT /*tag= c
FT /*number= 1
FT 231..3194
FT /*tag= d
FT /*number= 2
XX
XX US5874527-A.
XX PN
XX 23-FEB-1999.
XX PD
XX 30-SEP-1996; 96US-0719822.
XX PF
XX 02-JUN-1993; 93US-0072610.
XX PR
```


Db	1681	ataacgattaacgtatgcagacatatatacctctttggtgcgaattatttataacaataactca	1740
Qy	1741	aggaacctgatgcagagagcgctgaacagttaccattacaatcaaaaggaagcaccttgtacaagtac	1800
Db	1741	aggaacctgatgcagagagcgctgaacagttaccattacaatcaaaaggaagcaccttgtacaagtac	1800
Qy	1801	cagtgccagttagggcccgccaagaagtgcacaacggaagaatttgatgcactccaagagg	1860
Db	1801	cagtgccagttagggcccgccaagaagtgcacaacggaagaatttgatgcactccaagagg	1860
Qy	1861	acgatttcgaattagaagaacctgcagagctccagaggaaggaagatttagtattagaag	1920
Db	1861	acgatttcgaattagaagaacctgcagagctccagaggaaggaagatttagtattagaag	1920
Qy	1921	gagaaggagaaccaacgagaagaagccaagagagagagccaacagaaggagaagtgc	1980
Db	1921	gagaaggagaaccaacgagaagaagccaagagagagccaacagaaggagaagtgc	1980
Qy	1981	cagaagaagaatttagagccaactccagaggagcgatttcgaattagaagaaccaacaggag	2040
Db	1981	cagaagaagaatttagagccaactccagaggagcgatttcgaattagaagaaccaacaggag	2040
Qy	2041	aegaagttagaagaacccgtagaggcggaagaacctgcagaaggagagaagtggaagagg	2100
Db	2041	aegaagttagaagaacccgtagaggcggaagaacctgcagaaggagagaagtggaagagg	2100
Qy	2101	tacctgcagaagttagaagaagtgcgaagaggttacctgcagaagttagaagaagtggaagagg	2160
Db	2101	tacctgcagaagttagaagaagtgcgaagaggttacctgcagaagttagaagaagtggaagagg	2160
Qy	2161	taccagaagaagttagaagaagttaccgcgagaagttagaagaagtgcgaagaggttaccagaag	2220
Db	2161	taccagaagaagttagaagaagttaccgcgagaagttagaagaagtgcgaagaggttaccagaag	2220
Qy	2221	aegtgcagaagttaccagaagaagtgcgaagaggttaccagaagaagtggaagaggttaccag	2280
Db	2221	aegtgcagaagttaccagaagaagtgcgaagaggttaccagaagaagtggaagaggttaccag	2280
Qy	2281	aagaagtgcgaagaagtgcgaagaagttagaagaaggttagaagaagtggaagaggttaccagaag	2340
Db	2281	aagaagtgcgaagaagttagaagaaggttagaagaaggttagaagaagtggaagaggttaccagaag	2340
Qy	2341	aagtaccagcggttagtagaagaagaaggtgcacagaagaagttagaagaagaagaagaagagg	2400
Db	2341	aagtaccagcggttagtagaagaagaaggtgcacagaagaagttagaagaagaagaagaagagg	2400
Qy	2401	aagaaccagttagagaagaagaatgtattacaattagtaataccatccggaagaagaatatac	2460
Db	2401	aagaaccagttagagaagaagaatgtattacaattagtaataccatccggaagaagaatatac	2460
Qy	2461	aattagacaaccaagaagaagcaaattaggctctggaaattttatctatcatcgacatgc	2520
Db	2461	aattagacaaccaagaagaagcaaattaggctctggaaattttatctatcatcgacatgc	2520
Qy	2521	actacaagaagcttccaaagggaatttaggaagaagaagaagaacgtgcagtgatccat	2580
Db	2521	actacaagaagcttccaaagggaatttaggaagaagaagaagaacgtgcagtgatccat	2580
Qy	2581	tgaaccagaagaatttgcagaaggaagattcacatctcacagaatggtcctcattcatc	2640
Db	2581	tgaaccagaagaatttgcagaaggaagattcacatctcacagaatggtcctcattcatc	2640
Qy	2641	aaggccttagaagcgactgggaacgattagaagtgagcttaataaaggcttagagaagaagat	2700
Db	2641	aaggccttagaagcgactgggaacgattagaagtgagcttaataaaggcttagagaagaagat	2700
Qy	2701	ggatggaacaagaataaagaattggctggctggcttcgcttaattgaaaaataaattggt	2760
Db	2701	ggatggaacaagaataaagaattggctggctggcttcgcttaattgaaaaataaattggt	2760
Qy	2761	cagaataatagtcataattcaacaaagaagaaggacccagctcgttttgagaaaaacagagt	2820
Db	2761	cagaataatagtcataattcaacaaagaagaaggacccagctcgttttgagaaaaacagagt	2820

Qy	2821	ggagcagcagagaaatgggtttaaagcagaagtccaaatcccaaatgattcac	2880
Db	2821	ggagcagcagagaaatgggtttaaagcagaagtccaaatcccaaatgattcac	2880
Qy	2881	acttgaaaaaatggatgaacgcacactcattccaatttttaaataattctgtgaaagata	2940
Db	2881	acttgaaaaaatggatgaacgcacactcattccaatttttaaataattctgtgaaagata	2940
Qy	2941	tgtcaccaattgaaaaacaagaaaccaaagaatggttaatgaatcactggaaaaagaacg	3000
Db	2941	tgtcaccaattgaaaaacaagaaaccaaagaatggttaatgaatcactggaaaaagaacg	3000
Qy	3001	aacgggggttatggttctgaaatcatttgaaagtattgaccacatcaaaattattaaatgtgg	3060
Db	3001	aacgggggttatggttctgaaatcatttgaaagtattgaccacatcaaaattattaaatgtgg	3060
Qy	3061	ctaagagtcgagaatgtagtcctgccaatccttaataataatagagaagaagaactca	3120
Db	3061	ctaagagtcgagaatgtagtcctgccaatccttaataataatagagaagaagaactca	3120
Qy	3121	tgaatatggtttcctcaaaagaaaacgcaatatttaggcaagaatggaaaaaatggact	3180
Db	3121	tgaatatggtttcctcaaaagaaaacgcaatatttaggcaagaatggaaaaaatggact	3180
Qy	3181	cattggaaaaaagttaaatttttggttgccaattcaatggtacacaacattttctggaaaa	3240
Db	3181	cattggaaaaaagttaaatttttggttgccaattcaatggtacacaacattttctggaaaa	3240
Qy	3241	cgcttaaccaagaagaatggaaatcaatttggtaagaaataaaagtgttgaaattatgaa	3300
Db	3241	cgcttaaccaagaagaatggaaatcaatttggtaagaaataaaagtgttgaaattatgaa	3300
Qy	3301	aaaagaacagattattctctcttaataataataaattc	3337
Db	3301	aaaagaacagattattctctcttaataataataaattc	3337
RESULT 3			
AAH76457			
ID	AAH76457 standard; DNA; 3337 BP.		
XX			
AC	AAH76457;		
XX			
DT	22-OCT-2001 (first entry)		
XX			
DE	Plasmodium vivax ESP-1 DNA.		
XX			
KW	Plasmodium vivax; ESP-1; erythrocyte secreted protein-1; PvESP-1;		
KW	species-specific; malarial peptide antigen; infection; diagnosis;		
KW	malaria; ds.		
XX			
OS	Plasmodium vivax.		
XX			
FH	Key Location/Qualifiers		
FT	CDS 1..3197		
FT	/*tag= a		
FT	/*product= "ESP-1"		
FT	1..91		
FT	/*tag= b		
FT	/*number= 1		
FT	92..230		
FT	/*tag= c		
FT	/*number= 1		
FT	231..3197		
FT	/*tag= d		
FT	/*number= 2		
XX			
PN	US6231861-B1.		
XX			
PD	15-MAY-2001.		
XX			
PF	05-JUN-1998; 98US-0092458.		

Db 1621 agcctgtcaaaccaatggtgctccacgcgcagatgaaactttattctgttatcttag 1680
QY 1681 ataacgatttaacgtatgcagacattacatcctttgagcattatttaacaaatcctca 1740
Db 1681 ataacgatttaacgtatgcagacattacatcctttgagcattatttaacaaatcctca 1740
QY 1741 aggtacctgatgcagagaggtgttaacagttaccatcaaaagaaagcacctgttacaagtac 1800
Db 1741 aggtacctgatgcagagaggtgttaacagttaccatcaaaagaaagcacctgttacaagtac 1800
QY 1801 cagtgacgtaggccgcgcgcgaagaagtgccaacggaagaattgtgcaactccaagaag 1860
Db 1801 cagtgacgtaggccgcgcgcgaagaagtgccaacggaagaattgtgcaactccaagaag 1860
QY 1861 agcatttcgaattagaagaactgcgaagctccagaggaagagaaattagttattagaag 1920
Db 1861 agcatttcgaattagaagaactgcgaagctccagaggaagagaaattagttattagaag 1920
QY 1921 ggaaggaagaaccaaaccggaagaagcccaagagaagagagccaacagaaagagaagtgc 1980
Db 1921 ggaaggaagaaccaaaccggaagaagcccaagagaagagagccaacagaaagagaagtgc 1980
QY 1981 cagaagaagaattagaggaactccagaggaagcagatttcgaattagaagaaccaacagagag 2040
Db 1981 cagaagaagaattagaggaactccagaggaagcagatttcgaattagaagaaccaacagagag 2040
QY 2041 aagaagtagaagaacccgttagaggcggaagaactcagaagaagagaagtggaagaag 2100
Db 2041 aagaagtagaagaacccgttagaggcggaagaactcagaagaagagaagtggaagaag 2100
QY 2101 tacctgcagaagtagaagaagtggaaaggttacctgcagaagtagaagaagtggaaagag 2160
Db 2101 tacctgcagaagtagaagaagtggaaaggttacctgcagaagtagaagaagtggaaagag 2160
QY 2161 taccagaagaagtagaagaaggttaccgcgagaagtagaagaagtggaaaggttaccagaag 2220
Db 2161 taccagaagaagtagaagaaggttaccgcgagaagtagaagaagtggaaaggttaccagaag 2220
QY 2221 aegtgaaagaggtaccagaagaagtggaaaggttaccagaagaagtggaaaggtaccag 2280
Db 2221 aegtgaaagaggtaccagaagaagtggaaaggttaccagaagaagtggaaaggtaccag 2280
QY 2281 aagaagtgaaagaagtggaaagaagttagaagaagtagaggtaccagcggttagtagaagttag 2340
Db 2281 aagaagtgaaagaagtggaaagaagttagaagaagtagaggtaccagcggttagtagaagttag 2340
QY 2341 aagttaccagcggttagtagaagaagaaggttccagaagaagtagaagaagaagaagaag 2400
Db 2341 aagttaccagcggttagtagaagaagaaggttccagaagaagtagaagaagaagaagaag 2400
QY 2401 aagaaccagtagaggaagaagatgtattacaattagtaataccatcgggaagatatatc 2460
Db 2401 aagaaccagtagaggaagaagatgtattacaattagtaataccatcgggaagatatatc 2460
QY 2461 aattagacaaccaaagaagaacgaattaggctctggaatttttatctatcatcgacatgc 2520
Db 2461 aattagacaaccaaagaagaacgaattaggctctggaatttttatctatcatcgacatgc 2520
QY 2521 actaccaagaagcttccaaaggaattttatggagaagaagaagaacagtgagtgatccat 2580
Db 2521 actaccaagaagcttccaaaggaattttatggagaagaagaagaacagtgagtgatccat 2580
QY 2581 tgaaccagaagaattttgcaagaagaatttcacaattcagaatggctcacattcattc 2640
Db 2581 tgaaccagaagaattttgcaagaagaatttcacaattcagaatggctcacattcattc 2640
QY 2641 aaggcctagaagcgactgggaacgattagaagtgagcttaataaaggcttagagaagaat 2700
Db 2641 aaggcctagaagcgactgggaacgattagaagtgagcttaataaaggcttagagaagaat 2700
QY 2701 ggaatgaaacaagaataaagaatggctggctggctggctggctggctggctggctggctggct 2760
Db 2701 ggaatgaaacaagaataaagaatggctggctggctggctggctggctggctggctggctggct 2760

RESULT 4

AAT05868

ID AAT05868 standard; DNA; 3399 BP.

XX

AC AAT05868;

XX

DT 14-AUG-1996 (first entry)

XX

DE Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.

DE

KW Chicken leucocytozoan; immunogen; recombinant vaccine; protection;

KW

immunisation; vaccination; ss.

XX

OS Chicken leucocytozoan.

XX

FH Key Location/Qualifiers

FT CDS

1..3399

/*tag= a

1150..3218

/*tag= b

/note= "fragment referred to in the claims, for

use as insert in a recombinant vaccine

against chicken leucocytozoan disease"

XX

PN JP07284392-A.

XX

PD 31-OCT-1995.

XX

PF 19-APR-1994; 94JP-0080643.

XX

PR 19-APR-1994; 94JP-0080643.


```
QY 1191 ccagagatgatttcgattacacgaactgcagaagcagaaggaacacagaacacgtagag 1250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1206 cgaggagatgacgatgatgagcaaatgagcagcagcagcagcagcagcagcagcagca 1265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1251 ggaagaagaacacagaagaagctgcagaagcagaagcagcagcagcagcagcagcagcag 1310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1266 gaaggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1311 gaagagtagaggcacaactccagcagcagcagcagcagcagcagcagcagcagcagcagcag 1370
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1326 acagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1371 accgaagaactgcagaagcagaagcagcagcagcagcagcagcagcagcagcagcagcagcag 1427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1386 cctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1428 gaagaacacccctagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1487
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1446 gcaggagccacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1505
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1488 gagatgacttccaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1547
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1506 ggaagcacaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1565
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1548 ggggaagcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1607
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1566 ggaagcacaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1608 gtgactcctgctcagcctgcacaacaaatggtcgtcccaacgagcagcagcagcagcagcagcagcagcag 1667
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1626 ggaagcacaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1685
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1668 gttgatctcttagataacgattaaagctatgcagacattacatcctcttgagccattattt 1727
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1686 ggaagcacaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1745
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1728 aaacaaatcctcaaggatcctgtatgcaggagcagcagcagcagcagcagcagcagcagcagcagcagcag 1787
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1746 ggaagcacaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1805
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1788 cctgtacaagtaccagctgcagtagggccgcgcagagaagtgccaaagcagcagcagcagcagcagcagcagcag 1847
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1806 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1865
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1848 caactccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1907
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1866 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1925
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1908 ttgatttagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1967
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1926 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1985
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1968 gaaggagaagtgcagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2027
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1986 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2045
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2028 gaaccaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2087
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2046 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2088 gaagtgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2106 gcaggatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2148 gaagtgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2166 gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2208 gaggtaccagaagaagtgaagaggctaccagaagaagtgaagaggtagcagcagcagcagcagcagcagcagcag 2267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2226 gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

2268 gaagaggtaccagaagaagtggagaagtggagaagtgaagaagttagaggtaccagcg 2327
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

2286 ggagagcag 2345
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

2328 gtagtgaagtgaagtaccag 2387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

2346 gttagaggagcag 2405
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

2388 gaagaagaagcag 2447
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

2406 ggaagttagagcag 2465
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

2448 gaagaagaatatacaattag 2466
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

2466 gcaggagttagaggagcag 2484
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 16
AAF82901
ID AAF82901 standard; DNA; 3489 BP.
XX
AC AAF82901;
XX
DT 29-JUN-2001 (first entry)
XX
Nucleotide sequence of KSHV tethering protein, LANA.
XX
DE Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
XX KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
KW KSHV; latency-associated nuclear antigen; LANA; ds.
XX
OS Kaposi's sarcoma associated herpesvirus.
XX
FH Key
FT CDS 1..3489
FT /*tag- a
XX
PN WO200125484-A2.
XX
PD 12-APR-2001.
XX
PE 29-SEP-2000; 2000WO-US26908.
XX
PR 01-OCT-1999; 99US-0410399.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Robertson ES, Cotter MA;
XX
DR WPI; 2001-281736/29.
DR P-PSDB; AAB62331.
XX
PT A composition for use in gene therapy comprises an expression vector
PT that includes a nucleic acid sequence encoding a nucleic acid binding
PT protein -
XX
PS Disclosure; Fig 9A; 60pp; English.
XX
XX The invention provides a composition comprising nucleic acid, histone H1
CC protein and expression vector operationally encoding a protein suitable
CC for tethering the nucleic acid to the histone H1 protein, where the
CC tethering protein is LANA. The composition is useful in aiding the
CC retention of the viral DNA in the host cell. The viral vector encodes a
CC protein suitable for tethering DNA to Histone H1. Methods for screening
CC for compounds which are agonistic or antagonistic for the tethering of
CC viral proteins to histone H1 and DNA binding sites are useful for
CC developing the method of viral transfer. The composition has applications
CC to gene therapy, including the treatment of multiple sclerosis,
CC Parkinson's disease, Huntington disease and diabetes. The present
CC sequence represents the nucleotide sequence of the Kaposi's sarcoma
CC associated herpesvirus (human herpesvirus 8) latency-associated nuclear
CC antigen (LANA), which acts as a tethering protein.
XX

This sequence is a fragment of the Kaposi's sarcoma-associated herpesvirus (KSHV) LUR (long unique region). This fragment contains coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67 which encodes tegument protein IV, ORF68 which encodes a glycoprotein, ORF69, K12 which encodes kaposin, K13, ORF72 which encodes cyclin D, ORF73 which encodes immediate early protein (IEP), K14 which encodes OX-2 (v-adh), ORF74 which encodes G-protein coupled receptor, ORF75 which encodes tegument protein/FGFRAT, K15. KSHV is a new human Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the most common form of neoplasm occurring in persons with acquired immune deficiency syndrome (AIDS). The DHFR protein is useful for vaccination, prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma and for detecting expression of a DNA virus associated with Kaposi's sarcoma in a cell.

Db 125611 GGAGCCACAGCAGCGGAGCCACAGCAGCGGGAGCCACAGCAGCGGAGCCACAGCAGCG 125552
QY 1728 aaacaatctcgaagatcctgtagcagagagcgtgtaacagtagtaccatcaaaaggaagca 1787
Db 125551 GGAGCCACAGCAGCGGAGCCACAGCAGCGGGAGCCACAGCAGCGGAGCCACAGCAGCG 125492
QY 1788 cctgtacaagtaccagtgcagtaggcccgcgcgaagaagtcgcaacgcgaagaattgatg 1847
Db 125491 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 125432
QY 1848 caactccaagaggacgatttcgaattagaaaggaactgcgaagaactccagaggaagagaa 1907
Db 125431 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 125372
QY 1908 ttagtattagaaggaaggaacccaacgcgaagaagccaaagagagccaaagaggaagcaaca 1967
Db 125371 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 125312
QY 1968 gaagggaagtgccgaagaagaatttagaggcaactccagaggacgatttcgaattagaa 2027
Db 125311 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 125252
QY 2028 gaaccaacagggaagaagtagaagaacccgtagagggcgaaagaaactgcagaagagaa 2087
Db 125251 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 125192
QY 2088 gaagtggaaaggttacctcagaagtagaagaagtagaagtagaagtagaagtagaagtagaag 2147
Db 125191 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGTGA 125132
QY 2148 gaagtggaaaggttaccagaagaagtagaagaggtaccgcgagaagtagaagaagtggaa 2207
Db 125131 GCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 125072
QY 2208 gaagtaccagaagaagtgaagaggttaccagaagaagtagaagtagaagtagaagtagaagtg 2267
Db 125071 GCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 125012
QY 2268 gaagaggttaccagaagaagtcggaagagtggaagaagtagaagaagtagaagtagaagtagaag 2327
Db 125011 GGAGGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 124952
QY 2328 gtagtgaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaag 2387
Db 124951 GTTAGAGGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 124892
QY 2388 gaagaagaaggaagaaccagtagaggaagaagtagaagtagaagtagaagtagaagtagaagtagaag 2447
Db 124891 GGAGTTAGAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 124832
QY 2448 gaagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtag 2466
Db 124831 GCAGGAGTTAGAGGAGCAG 124813

RESULT 19
ABL05187/C
ID ABL05187 standard; cDNA; 4954 BP.
XX
AC ABL05187;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10043.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
XX

PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-556860/75.
DR P-PSDB; ABB61084.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 10043; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB957737-AB972072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4954 BP; 880 A; 1711 C; 1284 G; 1079 T; 0 other;

Query Match 5.4%; Score 181.8; DB 23; Length 4954;
Best Local Similarity 45.5%; Pred. No. 6e-21;
Matches 891; Conservative 0; Mismatches 1047; Indels 21; Gaps 6;
QY 499 cggaacaagaagagactacacctggtgatatgaagagagaagaagcctggcgatctgg 558
Db 3076 CAGCAGGAGCAGCAGCAGCAGGAGATCCGTCAGCTGGAGCGCCAGGAGCGCAGATG 3017
QY 559 aagcaggggaagagctggtgatttggaaagcaggggaagaactggcgttttggaaagcag 618
Db 3016 ATCCATCGGCTGGAGCAGCAGGAGCGCAGCTTGTGACATCAGCAGGAGCCAGGAGCAG 2957
QY 619 ggaagaagaactggcgatttggaaagcaggggaagaagcgtggttggaaagcaggggaag 678
Db 2956 CGGAAGATCCTTCAGCTGGAGAACCCAGGAGCGGCAGATGATCCGTCAGCTGGAGCGCCAG 2897
QY 679 aaactggcgtttggaagcaggggaagaagaactggagatcggaactgaagaaggaagca 738
Db 2896 GAGCGGCAGTTGTGACATCAGCAGGAGAGCCAGGAGCAGCGGAGATCCGTCAGCTGGAG 2837
QY 739 ctggagatcggaagaactgaaatggagcaactgtgtatgtagacacagaagaatagttcag 798
Db 2836 CGCCAGGAGCGGCAGTGTGACATCAGCAGGAGAGCGCAGGAGCAGCGAAGATCCGTCAG 2777
QY 799 ctgtaggagcagaaaaagtagatcttctcgtctcaaaaaatgtacaacctgcgtagta 858
Db 2776 CTGGAGCGCCAGGAGCGGCAGATGATCCATCCGCTGGAGCAGCAGGAGCGGCAGTGTGA 2717
QY 859 atgatgccctcttggaaagtattttgataaagataataattttgatcatattaaagatt 918
Db 2716 CATCAGCAGGAGAGCCAGGAGCAGCGGAAGATCCGTCAGCTGGAGCGCCAGGAGCGCAG 2657
QY 919 tcgagccactattcgaacaaattgtggcgggtactgtactgaacatgtttacgggacaagaat 978
Db 2656 ATGATCCATCGGCTGGAGCAGCAGGAGCGGCAGTTGTGACATCAGCAGGAGAGCAGGAG 2597
QY 979 tgccaatgaacctgtaccattaccagtggcagaagaagccgcgcaagtagtccagcggaag 1038

Db 2596 CAGCGGAAGATCCT-----TCAGCTGGAGAACACGAGCGCGCAGATGATCCATCGGCTG 2543
Qy 1039 aattagatgccactccagagatgaacttcgcattagattacagaaatcccccaggaag 1098
Db 2542 GAGCACCAGGAGCGGAGTTGTGACATCAGCAGGAGAGCCAGGAGCGCGGAAGATCCGT 2483
Qy 1099 tagaattagattagatgaagaggaactgaagaatacaacggaagtgggaccacgg 1158
Db 2482 CAGCATCAGCAGGAGAGCCAGGAGCGGGAAGATCCGTGAGTGGAGCGCCAGGAGCGG 2423
Qy 1159 aagaagaccacacgaagaattagatgccactccaagagatatttcgattagacga 1218
Db 2422 CAGATGATCCATCGCTGGAGCACACGAGCGCGCATGTTG-TGACATCAGCAGGAGAGCCCA 2364
Qy 1219 actgcagaaggaagaacacagaagaacgtagagggagaagaacacagaagaagctgcagaag 1278
Db 2363 GGAGCAGCGGAAGATCCGTGAG--CTGGAGCGCCAGGAGCGCGAGATGATCCATCGGCTG 2306
Qy 1279 gagaagtatcagaagaactccgaaggaagaagaagtttagagggaactccagagatg 1338
Db 2305 GAGCACCAGGAGCGGAGTTGTGACATCAGCAGGAGCGCGCAGGAGCGCGGAAGATCCCTT 2246
Qy 1339 atttcgattagatgaactacattagaagaacacgaagaactgcagaagggagaagaa 1398
Db 2245 CAGCTGAGAACACGAGCGCGCAGATGATCCATCGGCTGGAGCACACGAGGCGCGCATGTTG 2186
Qy 1399 ccgtagagggaagaacccgttagagggagaagaacacccgttagagggagaagaagctgcag 1458
Db 2185 TGACATCAGCAGGAGAGCCAGGAGCAGCGAGAGATCGTCACTGGAGCGCCAGGAGCGG 2126
Qy 1459 aaggagaagaagtttagaggaactccagagatgaactccattagaagaaccatcag 1518
Db 2125 GAGAAGATCCATCAGCTGGAGCGCCAGGAGCGCGCATGTTGTGACATCAGCAGGAGAGCCAG 2066
Qy 1519 gagaagagagaagggagaaggaagggagaaggaaggaagcgttagtagcag 1578
Db 2065 GAGCAGCGGAAGATCCGTGAGTGGAGCGCCAGGAGCGCGCATGTTGTGACATCAGCAGGAG 2006
Qy 1579 tgcagtagtggccgaaccggttagaagtagtgaactcctgcagcctgtcaaacccaatgg 1638
Db 2005 AGCCAGGAGCAGCGGAAGATCCGTGAGTGGAGCGCGCAGGAGCGCGCATGATCCATCGG 1946
Qy 1639 tcgctcaacgagcagatgaactttattcgttgcattcttagataacgatttaacgtatg 1998
Db 1945 CTGGAGCACCAGGAG-----CGGCAGTTGTGACATCAGCAGGAGCGCCAGGAGCAGCGG 1892
Qy 1699 cagacattacattctttgagcattatttaacaaatccctcaaggatcctcgtatgcaggag 1758
Db 1891 AAGATCCTTCAGCTGGAGAACACGAGCGCGCAGATGATCCATCGGCTGGAGCACCAGGAG 1832
Qy 1759 aggc---tgtaacagttaccatcaaaaggaagcactgtacaagtaccagtggcagtagggc 1815
Db 1831 CGGCAGTTGTGACATCAGCAGGAGCGCGCAGGAGCGCGCAGATGATCCATCAGCTGGAGCGC 1772
Qy 1816 ccgcgaagaagtgccaaacggaagatttgatgcaactccaaagagcagatttcgaattag 1875
Db 1771 CAGGAGCGGAGTTGTGACATCAGCAGGAGAGCAGCAGGAGCGCGGAAGATCCGTGAGCTG 1712
Qy 1876 aaggaactgcagaagctccagaggaaggaatttagattagaaggaaggaaggaaccaa 1935
Db 1711 GAGCGCAGGAGCGCGCAGATGATCCATCAGCTGGAGCGCCAGGAGCGCGCATGTTGTGACAT 1652
Qy 1936 cggaagaagacgaagaagagagccaaacagaaggaagtgccagaagaagaattag 1995
Db 1651 CAGCAGGAGAGCCAGGAGCAGCGGAAGATCCGTGAGTGGAGCGCCAGGAGCGCGCAGATG 1592
Qy 1996 agccaactccagaggaagcagtttcgaattagaagaacacagaagaagtagagaagaaa 2055
Db 1591 ATCCATCGGCTGGAGCACACGAGGAGCGCGCATGTTGTGACATCAGCAGGAGCGCCAGGAGCAG 1532
Qy 2056 ccgttagagggcgaagaactgcagaaggaagaagtggaagaggtaccctgcagaagtag 2115
Db 1531 CGGAGATCCTTCAG---CTGGAGAACACGAGCGCGCAGATGATCCATCGGCTGGAGCAC 1475

Qy 2116 aagaagtgaagaggtacctgcagaagtgaagaagtgaagaggtaccagaagaagtag 2175
Db 1474 CAGGAGCGGCAGTTGTGACATCAGCAGGAGAGCCAGGAGCGCGGAAGATCCGTGAGCTG 1415
Qy 2176 aagaggtaccgccagagtagaagaagtgaagaggtaccagaagaagtgaagaaggtac 2235
Db 1414 GAGCGCCAGAGCGGAGATGATCCATCGGCTGGAGCAGCAGGAGCGCGCATGTTGTGACAT 1355
Qy 2236 cagaagaagtgaagaggtaccagaagaagtgaagaggtaccagaagaagtgaagaag 2295
Db 1354 CAGCAGGAGCGCGCAGGAGCAGCGGAAGATCCTTCAGCTGGAGAACACGAGGAGCGCGCATG 1295
Qy 2296 tgaagaagtgaagaagttagaggtaccagcggttagaagtagaagtaccagcggttag 2355
Db 1294 ATCCATCGGCTGGAGCAGCAGGAGCGCGCATGTTGTGACATCAGCAGGAGAGCCAGGAGCAG 1235
Qy 2356 tagaagaagaagtgccagagaagaagtgaagaagaagaagaaggaaggaaccagtagagg 2415
Db 1234 CGGAAGATCCGTGAGTGGAGCGCCAGGAGCGCGCATGATCCATCGGCTGGAGCACCAG 1175
Qy 2416 aagaagaatgtattacaaattagtaataaccatcggaagaag 2454
Db 1174 GAGCGGCAGTTGTGACATCAGCAGGAGAGCCAGGAGCAG 1136

RESULT 20

ABL05186/C
ID ABL05186 standard; cDNA; 9516 BP.

XX ABL05186;

XX DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 10040.

DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB61083.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

XX Claim 1; SEQ ID NO 10040; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO


```
RESULT 21
AAQ03875
ID AAQ03875 standard; DNA; 3095 BP.
XX AC AAQ03875;
XX DT 24-AUG-1990 (first entry)
XX DE Sequence encoding carboxylic terminal part of native GLURP.
XX KW Plasmodium falciparum; antigen; malaria; vaccine; GLURP;ss
XX OS Plasmodium falciparum.
XX FH Key Location/Qualifiers
XX CDS 1..2352
XX FT /*tag= a
XX FT /product=GLURP
XX PN W09022811-A.
XX PD 22-MAR-1990.
XX PF 18-SEP-1989; 89WO-0000218.
XX PR 03-MAR-1989; 89US-0218885.
XX PR 03-MAR-1989; 89DK-0005191.
XX PA (STAT-) STATENS SERUMINST.
XX PI Driegiel M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A;
XX PI Jakobsen PH;
XX DR WPI: 1990-115998/15.
XX DR P-PSDB; AAR05804.
XX PT Polypeptide(s) derived from Plasmodium falciparum antigen - used in
XX PT vaccines and in production of antibodies, for diagnosis and
XX PT therapy of malaria.
XX PS Disclosure; Fig 7; 108pp; English.
XX CC An open reading frame of 2349 bps extends from the 5' terminal end of the
XX CC insert to a "TAA" stop codon. It is longest ORF found in the sequence.
XX CC Sequence displays some of the characteristics of other malaria nucleic
XX CC acid sequences : tandemly repeated motifs, high AT content and a
XX CC corresponding preference for codons containing these bases, and a high
XX CC content of codons for glutamate. Three major repetitive sequences are:
XX CC one motif from bp 34 to bp 156 is repeated from bp 289 to bp 411; another
XX CC motif from bp 477 to bp 521 is repeated tandemly twice from bp 522 to bp
XX CC 566 and from bp 567 to bp 611; a third motif from bp 1174 to bp 1233 is
XX CC repeated tandemly 11 times. This last repetitive region consists of 360bp
XX CC repeats differing only in 3 bases GAT coding for aspartate. This region
XX CC is flanked to the 5' terminal od a degenerated 60 bp repeat. GC content
XX CC of the coding part of the insert is on average 30%, and of the non-coding
XX CC 3' terminal 11%.
XX SQ Sequence 3095 BP; 1443 A; 300 C; 491 G; 861 T; 0 other;

Query Match 5.1%; Score 171.8; DB 11; Length 3095;
Best Local Similarity 44.8%; Pred. No. 2.4e-19;
Matches 846; Conservative 0; Mismatches 1022; Indels 21; Gaps 4;

QY 597 gaaactggcatttgaagcaggagggaagaactggcgatttgaagcaggagggaagaagct 656
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 16 gaaaaagcagcatgaagcagctgaaatgaagaagtgttgaagaagccatcat 75

QY 657 ggtgatttgaagcaggagggaagaactggcgatttgaagcaggagggaagaactggagat 716
    || || || || || || || || || || || || || || || || || || || ||
DB 76 gaagaaattgtacctgacacaaataaagaagaatcaggtgaagtaaatagttgataat 135

QY 717 gcggaaactgaagaaggagcaactggagatgcggaaactgaaatgagcaactgtgat 776
```

```
DB 136 gatgaaggtgttttgaagaagctcatcatgaaaaattttcatctgaagtaagtactct 195
QY 777 gttagacacagaagatgttcagctgtgagcagagaaaaagtacatgttctctgctcaagaa 836
    || || || || || || || || || || || || || || || || || || || ||
DB 196 gaattaaatgaagaatgaatttgtgaatctgacaaaagtgttaactgaacctgtcgaacat 255
QY 837 aatgtacaacctccgataagtaatgagtcctcttttggaaagtattttggataaagatatata 896
    || || || || || || || || || || || || || || || || || || || ||
DB 256 gaagaa-----gttgtatctgaagaagcaacctgaacctgaaaaatgaagaagt 309
QY 897 atttttgatcatattaaagatttcgagccactatttgcacaaatttggcgggtactgct 956
    || || || || || || || || || || || || || || || || || || || ||
DB 310 agtatagaagaagctcatcaggaagaattgtacctgaaacaaatgatgaagaatcaggt 369
QY 957 aaacatgtttacgggacaagaattgccaatgaaacctgtaccattaccagtggcagaagag 1016
    || || || || || || || || || || || || || || || || || || || ||
DB 370 gaaagtggattagatgaatgaagaagtgatttggagaacctaatcatgaagaattt 429
QY 1017 ccgcgcgaagtaccagcggaagaattagatgccactccagaggtgacttgcgcattagat 1076
    || || || || || || || || || || || || || || || || || || || ||
DB 430 gaacctgatcaaaatgactctgaattaaatgaagaatgaattagtgaaacaaaaagt 489
QY 1077 gttacagaatctcccgagggaagttagaattagttattagatgaagagccaactgaagaagaa 1136
    || || || || || || || || || || || || || || || || || || || ||
DB 490 gtatctgaaccagctgaacatgtagaattgtatcagaaaaaagtgtatctgaacctgact 549
QY 1137 tcaacggaagtggaccaacgcgaagagcacaacgcgaagaattagatgccactccagag 1196
    || || || || || || || || || || || || || || || || || || || ||
DB 550 gaacacgtgaaattgtatctgaaaaaagtacatccgaa-----ccagctgaaca 599
QY 1197 gatggatttcgcatttagacgaactgcagaagagagaacagaagaacctagaggggagaa 1256
    || || || || || || || || || || || || || || || || || || || ||
DB 600 tgtagaagtgtatctgaacaaagtaataacgaacctccga--aaagaagaatggacca 657
QY 1257 gaaacgaagaactgcagaagagagaagtatcagaagaacactccagaagagagaagaagag 1316
    || || || || || || || || || || || || || || || || || || || ||
DB 658 gtcctctcaaaacatttgaagaattgaaaagtggatttcaacctaaaaattgtgac 717
QY 1317 ttagaggcaactccagaggtgatttcgcatttagatggaaactacattagagaacacgaa 1376
    || || || || || || || || || || || || || || || || || || || ||
DB 718 ctcaataaattgaactaatatttgtgactcaacaaatccacaagaacctggtgaa 777
QY 1377 gaaactgcagaaggaagaacacgttagagggagaagaacaccttagagggagaagaaccc 1436
    || || || || || || || || || || || || || || || || || || || ||
DB 778 ccatctttgtcaaaattgaaaagtctcttctgagaagaaataaacatgcaagtgtgat 837
QY 1437 gttaggggagaagaactgcagaaggaagaagattagaggcaactccagaggatgac 1496
    || || || || || || || || || || || || || || || || || || || ||
DB 838 cctgaagtataaagaaaaaagaaatgtatctgaagtgttgaagaaaaacaaaattcaca 897
QY 1497 ttccaatagaagaacccatcaggagaaggggaaggggaaggaaggaaggggaagga 1556
    || || || || || || || || || || || || || || || || || || || ||
DB 898 gaatcagttgaacaaatccagtaaatgagatgaatttgaagatgttcacactgaacaa 957
QY 1557 gaaggagaagcgttagtagcagtgccagtagtgccgagacccggtagaagtagtgaactct 1616
    || || || || || || || || || || || || || || || || || || || ||
DB 958 ttagatttagatcataaaacagttgtaccagaaatagtagaagttgaagaattcctca 1017
QY 1617 gctcacctgtcaaaccaatgttcgctcccaacgcgcagatgaaactttatctgtgatc 1676
    || || || || || || || || || || || || || || || || || || || ||
DB 1018 gaactacatgaataatgaagtggctcatcccgaaatgttgaatttgagggaagtttctcct 1077
QY 1677 ttgataaacgatttaacgtatgcagacattacatctctttagagcattattttaacaacatc 1736
    || || || || || || || || || || || || || || || || || || || ||
DB 1078 gaaccaaatacaataacgaatttcaagaataataatgaagatgataaaagtgcacatatt 1137
QY 1737 ctcaagatcctgatgcaggagagaggtgttaacagttaccatcaaaagggaagcacctgtacaa 1796
    || || || || || || || || || || || || || || || || || || || ||
DB 1138 cagcatgaaatagtagaagtagaagaataactccagaagatgataaaatgaaaaagtt 1197
QY 1797 gtaccagtgagtagggcccgccgcaagaagtgccaacgcgaagaattgtgcaactccaa 1856
    || || || || || || || || || || || || || || || || || || || ||
```

Db 1198 gaacatgaaatagtagaagttgaagaaattctagcagaagataaaatgaaaaaggtcaa 1257
 QY 1857 gagagcatttcgaatagaaagaaactgcagaagctccacagaggaaggaattagatta 1916
 Db 1258 catgaaatagtagaggtgaagaattctaccagaagatgataaaatgaaaaaggtcaa 1317
 QY 1917 gaagggagaagagacacacggaaggaagagcgaaggaaggaaggaaggaagga 1976
 Db 1318 catgaaatagtagaagttgaagaattctaccagaagataaaatgaaaaaggtcaacat 1377
 QY 1977 gtccagaagaagaattagaggcaactccagagagcagatttcgaattagaagaaccaaca 2036
 Db 1378 gaattgtagaggttagaagaattctaccagaagataaaatgaaaaaggtgaaactgaa 1437
 QY 2037 ggagaagaagtagaagaacccgtagaggcggaagaactgcagaag--agaagaagt 2093
 Db 1438 atagtgaagttgaagaattctaccagaagataaaatgaaaaaggtcaacatgaata 1497
 QY 2094 gaagaggtacctcagaagtagaagaagtggaagaggtacctcagaagtagaagaagt 2153
 Db 1498 gtgaggttgaagaattctaccagaagataaaatgaaaaaggttcaacatgaataagta 1557
 QY 2154 gaagaggtaccagaagaagtagaaggttaccgcagaagtagaagaagtggaagagta 2213
 Db 1558 gaagttgaagaattctaccagaagataaaatgaaaaaggttcaacatgaatagtag 1617
 QY 2214 ccagaagaagtggaaggttaccagaagaagtggaagaggtaccagaagaagtggaagag 2273
 Db 1618 gtgagaagaattctaccagaagaagataaaatgaaaaaggttcaacatgaatagtag 1677
 QY 2274 gtaccagaagaagtggaagaagtggaagaagtagaagaagtagaggtaccagcggtagta 2333
 Db 1678 gtgagaagaattctaccagaagataaaatgaaaaaggttcaacatgaatagtagaggt 1737
 QY 2334 gaagtagaagtagcagcggtagtagaagaagaaggtgccaagaagtagaagaagaaga 2393
 Db 1738 gaagaattctaccagaagaataaaatgaaaaaggttcaacatgaatagtagaggtgaa 1797
 QY 2394 gaagaggaagaaccagtagagggaagaagtagtattcaattagataaccatcggaaga 2453
 Db 1798 gaattcttcagaagaattgtgaattgaagaagtagtaccatcaacaacaacaataatgaa 1857
 QY 2454 gatatacaattagacaacaacacagaaga 2482
 Db 1858 aatttgaaactataaacacacagaagaaa 1886

RESULT 22

AAT78868

ID AAT78868 standard; cDNA; 5361 BP.

XX AC AAT78868;

XX CT 08-OCT-1997 (first entry)

XX DE P. falciparum liver stage antigen-3 coding sequence.

XX KW Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;
 KW prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;
 KW glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;
 KW vaccine; immunotherapy; malaria; ds.
 XX OS Plasmodium falciparum.
 XX PN WO9641877-A2.
 XX PD 27-DEC-1996.
 XX PF 12-JUN-1996; 96WO-FR00894.
 XX PR 13-JUN-1995; 95FR-0007007.
 XX PA (INSP) INST PASTEUR.

XX Daubersies P, Druilhé P;
 XX WPI: 1997-065464/06.
 DR P-PSDB; AAW24790.
 XX Plasmodium falciparum poly:peptide(s) and related nucleic acids -
 PT derived from the liver stage antigen-3, useful for malaria vaccine
 PT prodn. and diagnosis
 XX Claim 20; Fig 2A-I; 69pp; French.

CC This sequence corresponds to the coding sequence for a Plasmodium
 CC falciparum strain K1 pre-erythrocytic liver stage antigen-3 (LSA-3)
 CC protein. The gene sequence was isolated by screening a P. falciparum
 CC strain T9/96 library with serum from a missionary treated by prophylaxis
 CC (for strain T6/96 see FR9101286). Of 20 clones isolated, clone 729S was
 CC used to screen a library generated from Thai strain K1. One clone
 CC contained a 6.85 kb insert including the genomic sequence AAT78867.
 CC The gene comprises a 1.8 kb region encoding 3 major blocks of 4 amino
 CC acid repeats and a 3' hydrophobic region corresponding to a
 CC glycosyl-phosphatidylinositol membrane anchoring sequence. The
 CC invention relates to new polypeptides of at least 10 amino acids derived
 CC from the LSA-3 polypeptide with the exception of the peptides AAW24791-4.
 CC The LSA-3 peptides can be used to raise antibodies and as vaccines for
 CC immunotherapy of malaria.

XX SQ Sequence 5361 BP; 2388 A; 431 C; 1169 G; 1373 T; 0 other;

Query Match 5.1%; Score 170.2; DB 18; Length 5361;
 Best Local Similarity 44.1%; Pred. NO. 4.9e-19;
 Matches 911; Conservative 0; Mismatches 1138; Indels 18; Gaps 4;

QY 407 tgaagaagtcacagaaacgttagatgaggtgaatcatcattagagaggttccactgaaga 466
 Db 306 tgaagaagtcacagaaacgttagatgaggtgaatcatcattagagaggttccactgaaga 365
 QY 467 tatggaaacagagagatggctcaacagatgatcggaacacagaaagaggaactcgtgta 526
 Db 366 ttagatgataataaaatttagaagaagccgaagtataaagaataatcttattaaag 425
 QY 527 tatggaagagagaagaagcgtggcgatattggaagcagggggaagagcgtggtattgga 586
 Db 426 taatataagaagaacacaaagaaatattattgacaattttattataataattggacaaa 485
 QY 587 agcagggggaagaactggcgatttgaagcagggggaagaacactggcgatttggagcagg 646
 Db 486 ttcagaaaaacacagaagaagtgtatcagaaaatgtacaagtcatgtgaaactttttaa 545
 QY 647 ggaagaagcgtggtgatttgaagcagggggaagaacactggcgatttggagcaggggaaga 706
 Db 546 attattaaatagttagatgtaattggaagaagtaaaagaataatttttggaggaagtca 605
 QY 707 aactgagatgcggaaactgaagaaggaagcaactggagatgcggaagaactgaataatggagc 766
 Db 606 agttaatgacgatatttttaattagtttagtaaaagtgttcaacaagaacaacaacaa 665
 QY 767 aactgtgtatgtacacagaagaagtagttcagctgtatggagcagaaaaatcatatttcc 826
 Db 666 tgttgaagaaaaagttgaagaaggtgtagaagaaaaatgcaggaagaagttgagaagaaa 725
 QY 827 tgcctcaagaaaaatgtacaacctgccgatagtaataatgacccctctttggagtagttttgga 886
 Db 726 tgtagaagaaaaatgtaga-----gaaatgacgcaggaagtgtgagcctcaagtgtga 779
 QY 887 taaagataataatttttgatcatatataaagatttcgagccactatttcgacaaatattgtgc 946
 Db 780 agaaagttagcttcaagtgttgataaagtatagattcaagtattgagaagaaatgtagc 839
 QY 947 gggtagctgctaaacatgttacgggacagaagaattgcacaatgaaacctgtaccattaccagt 1006
 Db 840 tccaactgttgaagaaatcgtagctccaaagtgt---tgtagaagttgtgctccaagtgt 896

QY 2144 agaagaagtgaagaggtaccagaagaagttagaagaggtaccgcgagaagttagaagaagt 2203
Db 2275 aagtgtgctcccaagtgttgaagaagttagaagaataattgaagaagttagctga 2334
QY 2204 ggaagaggtaccagaagaagtgaagaggtaccagaagaagtgaagaggtaccagaaga 2263
Db 2335 aaatgttgaagaagttagctgaaatgttgaagaagttagctgaaatgttgaaga 2394
QY 2264 agtgaagaggtaccagaagaagtgaagaagtgaagaagtgaagaagttagctacc 2323
Db 2395 aagtgtagctgaaatgttgaagaatacttagctccaaactgttgaagaatacttagctcc 2454
QY 2324 agcgttagtgaagaagttagctaccagcgttagtgaagaagaaggtgcccagaagaagttaga 2383
Db 2455 aactgttgaagaatacttagctccaaactgttgaagaagttagctccaaactgttgaaga 2514
QY 2384 agaagaagaagaaggaagaaccagtagaggaagaaggtattacaattagtaataacc 2443
Db 2515 aagtgtgaagaatactgttgaagaactgttagctgaaatgttgaagaagttagctga 2574
QY 2444 atcggaagaagaatatatacaattagacaa 2470
Db 2575 aaatgttgaagaagttagctgaaaa 2601

RESULT 24

AAZ23891
ID AAZ23891 standard; DNA; 49999 BP.
XX AC AAZ23891;
XX DT 25-JAN-2000 (first entry)
XX DE Murine LOBO genomic DNA fragment 1.
XX KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX OS Mus musculus.
XX PN WO9950284-A2.
XX PD 07-OCT-1999.
XX PF 26-MAR-1999; 99WO-EP02055.
XX PR 27-MAR-1998; 98DE-1013799.
XX PA (ROSE/) ROSENTHAL A.
XX PI Rosenthal A, Rump A, Hess J, Aligner T, Wirth T;
XX WPI; 1999-601320/51.
XX PT Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX PS Example 3; Page 69-97; 391pp; German.
XX CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX SQ Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;

Query Match 5.1%; Score 169.4; DB 20; Length 49999;
Best Local Similarity 66.2%; Pred. No. 9.5e-19;
Matches 260; Conservative 0; Mismatches 131; Indels 2; Gaps 1;
QY 2034 acagggaagaagttagaagaaccgttagagggcgagaagaactgcagaaggaagaagt 2093
Db 7380 atagaagaagaagaagagggagggagggaggaaggaaggaaggaagaagaag 7439
QY 2094 gaagaagctacctcagaagaagttagaagaagtgaagaaggtacctcagaagaagt 2153
Db 7440 gaagaagaaggaagaagaagaaggaaggaagaagaagaagaagaagaagaaga 7499
QY 2154 gaagaggtaccagaagaagttagaagaggtaccgcgagaagttagaagaagtgaagaagt 2213
Db 7500 gaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 7559
QY 2214 ccgagaagagtggaagaggtaccagaagaagtgaagaaggtaccagaagaagtgaaga 2272
Db 7560 gaagaagacagagagagggagggagggaggggggggaggaagaagaagaagaaga 7619
QY 2273 -ggtaccagaagaagtgaagaagtgaagaagttagaagaagttagaccagcggtag 2331
Db 7620 ggagagcgagaggaagaagaaggaggaagaagaagaagaagaagaagaagaaga 7679
QY 2332 tagaagttagaagttaccagcggttagtagaagaagaaggtgccagaagaagttagaagaaga 2391
Db 7680 gagaaggaagaagaagggagggagggaggaagaagaagaagaagaagaagaagaaga 7739
QY 2392 aagaagaagaagaaccagtagtagaagaagaagt 2424
Db 7740 aagaagaagaagaagagggagggagggaggg 7772
RESULT 25
AAZ23896
ID AAZ23896 standard; DNA; 49999 BP.
XX AC AAZ23896;
XX DT 25-JAN-2000 (first entry)
XX DE Murine LOBO homologue genomic DNA fragment 2.
XX KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX OS Mus musculus.
XX PN WO9950284-A2.
XX PD 07-OCT-1999.
XX PF 26-MAR-1999; 99WO-EP02055.
XX PR 27-MAR-1998; 98DE-1013799.
XX PA (ROSE/) ROSENTHAL A.
XX PI Rosenthal A, Rump A, Hess J, Aligner T, Wirth T;
XX WPI; 1999-601320/51.
XX PT Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX PS Example 3; Page 161-189; 391pp; German.
XX CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.

development), with exception of the skull and have osteopathic activity. The nucleic acid molecules, proteins and antibodies can be used in diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods and nucleic acid molecules etc. are useful for production of transgenic animals, especially a transgenic mouse for the study of diseases associated with bone development, e.g. spondyloepiphyseal dysplasia and achondroplasia. This sequence encodes the murine LOBO protein described in the method of the invention.

Query Match 5.1%; Score 169.4; DB 20; Length 49999;
Best Local Similarity 66.2%; Pred. No. 9.5e-19;
Matches 260; Conservative 0; Mismatches 131; Indels 2; Gaps 1;

[illegible]

RESULT	26
AAN60472	
ID	AAN60472 standard; DNA; 4590 BP.
XX	
XX	
AC	AAN60472;
XX	
XX	
DT	24-AUG-1991 (first entry)
XX	
DE	Sequence encoding the ring-infected Erythrocyte Surface Antigen (RESA).
XX	
KW	Malaria vaccine; antigen; epitope; ss.
XX	
OS	Plasmodium falciparum.

PR 11-SEP-1984; 84AU-0007066.
PR 10-SEP-1985; 85AU-0047326.
XX
XX
PA (HALL-) HALL INST MED RES.
XX
XX Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
PI WPI: 1986-094065/14.
XX DR P-PSDB; AAP60569.
XX
XX DNA coding for Plasmodium falciparum antigens - expressing
PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
PT falciparum
XX
XX Claim 4; Fig 1; 55pp; English.
PS
XX
XX The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base
CC sequence coding RESA (AAN60472) or FIRA (AAN60473). RESA and FIRA have
CC antigenicity suitable for providing protective immunity against
CC Plasmodium falciparum malarial infections.
XX
XX Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T; 0 other:
SQ

Query Match	5.1%	Score 168.8;	DB 7;	Length 4590;
Best Local Similarity	53.4%;	Pred. No. 8.1e-19;		
Matches 378;	Conservative 0;	Mismatches 327;	Indels 3;	Gaps 1;
QY	1845	atgcaactccaaaggagcatttcgaattagaagggaactgcagaagctccagaaggaagga	1904	
Db	3632	atgaaaacacaaatgaaatgtaccagaacatgtacacataatgctggaagaaaatgta	3691	
QY	1905	gaattagattagaaggagagagaccacaggagagccaaagagccaaagagagagcca	1964	
Db	3692	gaacatgatctgaagaaaatgtagaacatgatgctgaagaaaatgtagaacatgatgct	3751	
QY	1965	acagaaggagaatgcccagaagaagaattagggcgaactccagagacgatttcgaatta	2024	
Db	3752	gaagaaaatgtagaacatgatgctgaagaaaatgtagaacatgatgctgaagaaaatgta	3811	
QY	2025	gaagaaccaacagagaagaagttagaagaaacgttagagggcgaagaaactgcagaagga	2084	
Db	3812	gaagaaaatgtgaagaagttagaagaaaatgtgaagaaaatgtagaagaaaatgtaga	3871	
QY	2085	gaagaagtggagaaggtcacctgcagaagttagaagaagtggagaaggtcacctgcagaagt	2144	
Db	3872	gaaaatgttgaagaagttagaagaaaatgtagaagaaaatgtagaagaaaatgtagaagaa	3931	
QY	2145	gaagaagtggagaaggttaccagaagaagttagaagaggttaccgcagaagttagaagaagt	2204	
Db	3932	aatgttgaagaaaatgttgaagaaaatgttgaagaaaatgtagaagaaaatgtagaagaa	3991	
QY	2205	gaagaagttaccagaagaagtggagaaggttaccagaagaagtggagaaggttaccagaagaa	2264	
Db	3992	aatgttgaagaaatgatgtgaagaaaatgttgaagaaagttagaagaaaatgtagaagaaaat	4051	
QY	2265	gtggaaag--aggtaccagaagaagtgggaagaaggtggagaagatagaagaagtagagagta	2321	
Db	4052	gtagaagaaaatgtgaagaaaatgtagaagaaaatgttgaagaagttagaagaaaatgta	4111	
QY	2322	ccagcggtagtagaagttagcagcggtagtagaagaagaaggtgccagaagaagta	2381	
Db	4112	gaagaaaatgtgaagaaaatgtagaagagaatgttgaagagaatgttgaagagaatgta	4171	
QY	2382	gaagagaagaagaagaagggaacacagttaggggaagaagatgtattcaaatagtaata	2441	
Db	4172	gaagaatgatgaagaaaatgttgaagaacacaaatgaagaatatgatgaataaaaaaa	4231	
QY	2442	ccatcggagaagatatataattagacaaaccaaagaagaacgaattagtgcttggaaatt	2501	
Db	4232	atatatatataaagtttaattttataaacaggaataactactaaaagacagatttctct	4291	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 10:11:49 ; Search time 4066.1 Seconds
(without alignments)
17174.160 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gaattccggtaaagtaacaa.....tcttataaaataaataattc 3337

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797556 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description

1	3337	100.0	3337	6	ARI51085
2	3337	100.0	3337	6	I23337
3	342.8	10.3	2069	23	E10125
4	342.8	10.3	3399	23	E10126
5	330.6	9.9	11354	8	AP002460
6	318.6	9.5	81019	8	AB073159
7	307.8	9.2	144505	2	AL662884
8	307.2	9.2	1902	8	AF413055
9	304	9.1	215046	2	AC011767
10	299	9.0	172307	2	AC044842
11	294.6	8.8	142277	2	AC016204
12	283.6	8.5	16377	3	AF202180
13	283.6	8.5	102195	2	AC091338
14	278.8	8.4	272545	2	AC090533
15	278.4	8.3	174241	9	AC018634
16	277.4	8.3	282611	2	AL645746
17	273.8	8.2	165077	10	AC084382
18	268	8.0	1686	23	E08995
19	263.8	7.9	4630	3	PF111A
20	256.2	7.7	5420	3	AF056936
21	255.8	7.7	174019	2	AC027267
22	253.2	7.6	170276	2	AC106251
23	252.2	7.6	199508	2	AC098555
24	252	7.6	13297	10	AF311727
25	247.2	7.4	179149	9	AC009785
26	244.8	7.3	157158	2	AC098094
27	242	7.3	132449	9	AL365272
28	242	7.3	178022	2	AC018348
29	242	7.3	179337	2	AC091304
30	241.8	7.2	206136	9	AC068037
31	240	7.2	184855	2	AL606744
32	238.8	7.2	185497	2	AC073446
33	237.2	7.1	159971	2	AC091951
34	236.2	7.1	22398	5	FRU271723
35	235.6	7.1	208006	2	AC087877
36	234.8	7.0	3720	14	S76368
37	234.8	7.0	43658	14	HSV3PRGEN
38	234.8	7.0	112930	14	HSGEND
39	232.8	7.0	214864	9	AC069544
40	231.4	6.9	180385	9	AC007461
41	231.4	6.9	230339	10	AC068561
42	230	6.9	227605	2	AC073564
43	228	6.8	12029	3	AE001373
44	227.4	6.8	185994	2	AC002042
45	226	6.8	163542	2	AC106201

ALIGNMENTS

RESULT	1				
LOCUS	ARI51085	ARI51085	Sequence 1 from patent US 6231861.	3337 bp	DNA
DEFINITION	ARI51085	ARI51085	Sequence 1 from patent US 6231861.		
ACCESSION	ARI51085	ARI51085			
VERSION	ARI51085.1	GI:151117135			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 3337)				
AUTHORS	Barnwell,J.W.				
TITLE	Plasmodium vivax blood stage antigens, antibodies, and diagnostic assays				
JOURNAL	Patent: US 6231861-A 1 15-MAY-2001;				
FEATURES	Location/Qualifiers				
source	1..3337				
	/organism="unknown"				
BASE COUNT	1304 a	467 c	875 g	691 t	
ORIGIN					
Query Match			100.0%;	Score 3337;	DB 6; Length 3337;

Best Local Similarity 100.0%; Pred. No. 0;									
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	gaattccggttaaaagtaaacactatggttcgtatctatataataaccttactaatatttate	60						
Db	1	GAATTCGGTAAAGTAACAACTATGTGTTGCGTATCTATATATAACCTTACTAATTTATC	60						
QY	61	ttttgctttcttttaattcatctctcaacagtaagataaaaaataatctataaaaactgc	120						
Db	61	TTTTTGCTTTTCTTTTAATTCATCCTTCACAGTAAAGTAAAAATAATCTATAAAACTGC	120						
QY	121	tatatatacatatatttcataaagtggcatttgtgaattgcgatctatttaatttcgta	180						
Db	121	TATATATACATATATATTCATAAGTGGCATTTGTGAATTCGGATCATTTAAATTTACGTA	180						
QY	181	aaacaataattgaaaaaatt	240						
Db	181	AAAAAATAATTGAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTCACAGACGATTTAG	240						
QY	241	aattgaaaaatgcttctgatgtttagaggtgagatccctcaaacgcggttttag	300						
Db	241	AATTGAAAAATGCTTCTGATGATGTTGTAGAGGTGGAGGATCCTTCAAACGCGGTTTAG	300						
QY	301	aattagaagagaaaatttttgatgagaattcaagttgatgatgaactcttttagatgcta	360						
Db	301	AATTAGAAGAGGAAAAATTTTGATGGAATTCAGGTGATGATGAATCTTTTATAGATGCTA	360						
QY	361	ccccgaagatgacttgccttaacagatttgccaatgaaacacgatgaggaagtcaacg	420						
Db	361	CCCCGAGATGACTTTGCTTTAACAGATTTGCCAATTTGAACAGATGAGGAAGTCAACG	420						
QY	421	aaactgttagtgaggtgaatcaattaggagaggtttccactgaagatatggaacagaag	480						
Db	421	AAACGTTAGATGAGGTGAATCATATTAGGAGAGCTTTCCACTGAAGATATGGAACAGAAG	480						
QY	481	atgctcaacagatgatacggaaaacagaagaagactaactcgtgatatggaaggagaag	540						
Db	481	ATGGCTCAACAGATGATACGGAAACAGAGAAGGACTACTGCTGATATGGAAGGAGAAG	540						
QY	541	aagaagctggcgtatggaagcgggggaagaagcttggtgatttggaaagcaggggaagaa	600						
Db	541	AAGAAGCTGCGATATGGAAGCAGGGGAAGAAGCTGGTGNATTGGAAGCAGGGGAAGAAA	600						
QY	601	ctggcatttggaaagcaggggaagaactggcagatttggaaagcaggggaagaagctggcg	660						
Db	601	CTGGCATTTTGGAAAGCAGGGGAAGAACTGGCGATTTTGGAAAGCAGGGGAAGAGCTGGTG	660						
QY	661	atttgaagcaggggaagaaactgcgatttgaacaggggaagaagaactggaagatgcgg	720						
Db	661	ATTTGGNAGCAGGGGAAGAAACTTGGCGATTTTGGAACGGGGGAAGAACTGGAGATGCCG	720						
QY	721	aaactgaagaagcagcaactggagatgcggaactgaaactggagcaactgtgtatgtag	780						
Db	721	AAACTGAAGAAGCAGCAACTGGAGATGCGGAAACTGAAATGAGCAACTGTGTATGTAG	780						
QY	781	acacagaagatagttcagctgattgagcagaaaaagtaactgttctcgtctcaagaanaatg	840						
Db	781	ACACAGAAGATAGTTCAGCTGATGGAGCAGAAAAAGTACATGTTCTGTCTCAAGAAAAATG	840						
QY	841	tacaactcgcgatagtaaatgactcctcttggaaagtatttggataaagataataatt	900						
Db	841	TACAACCTGCCGATAGTAAATGATGCCCTCTTTTGGAAAGTATTTTGGATAAAGATATAATTT	900						
QY	901	ttgatcataattaagaatttcagccactattcgaacaaaatttggcggggtactgctaaac	960						
Db	901	TTGATCATATTAAGATTTTCAGCCACTATTTCGAACAAATTTGTGGCGGTACTGCTAAAC	960						
QY	961	atgttcagggaagaatttgcgaatgaacactgtaccattaccagttggcagaagaccgg	1020						
Db	961	ATGTTACGGGACAGAATTTGCCAATGAACCTGTACCATTTACCAGTTGGCAGAGAGCCCG	1020						
QY	1021	cgaagttaccagcggaagattagatgccactccagagatgaacttcgattagatgtta	1080						
Db	1021	TACCTCCAGAGTAGAAGAGTGGAAAGAGTACCTCTCCAGAGTAGAAGAGTGGAGAGG	1080						
QY	1081	cagaactctccagagaagtagaattagtagataaagcaactgaagaagaatcaa	1140						
Db	1081	CAGAATCTCCGAGGAAGTAGAATTAGTATTAGTAGAAGGCACTGAAGAAATCAA	1140						
QY	1141	cgaagtgggaccacgcgaaggaaggaaccacccaagaataattagatgccactccagagatg	1200						
Db	1141	CGGAAGTGGACCAACGGAAGAGGACCAACCGAAGAAATTAGATGCCACTCCAGAGGATG	1200						
QY	1201	gatttcgacttagaagaactcagaagagaagaacagaagaacgtagaggagaagaaa	1260						
Db	1201	GATTTTCGCATTAGCGAAACTTCAGAAAGGAGAAACAGAAACGTTAGGGGAGAAAGAAA	1260						
QY	1261	cagaagaactcagaagagaaglatcagaagaactccagaagagaagaagagcttag	1320						
Db	1261	CAGAAGAACTTCAGAAAGGAGAAAGTATCAGNAGNAACTCCAGNAGGAGAAAGAGATTAG	1320						
QY	1321	aggcaactccagagatgatttcgcaattagatggaaactacattagagaacacccaagaaa	1380						
Db	1321	AGGCAACTCCAGAGGATGATTTTCGCATTAGATGGAACCTACATTAGAAGAAACCGAAGAAA	1380						
QY	1381	ctgcagaagaagaagaacccgtagaggagaagaacccgtagaggagaagaacccgtag	1440						
Db	1381	CTGCAGAAGGAGAAACCCGTAGAGGGAGAAACCCGTAGAGGGAGAAACCCGTAG	1440						
QY	1441	agggagaagaagctgcagaaggaagaagagttagaggcaactccagaggaatgacttcc	1500						
Db	1441	AGGGAGAACAAGCTGCAGAAGGAGAGAAAGCTTACAGGCAACTCCAGAGGATGACTTCC	1500						
QY	1501	aattagaacacccatcaggagaaggaaggggaaggaaggaaggaaggaaggaaggaag	1560						
Db	1501	AATTAGAAGAACCATCAGGAGAGGAGAGGGGAGAGAGAGAGAGAGGAGAGGAGAAG	1560						
QY	1561	gagaagcgttagtagcagtgccagtagtgccgaacccgtagaagtagtgcctccgctc	1620						
Db	1561	GAGAAGCGTTAGTAGCAGTGGCCAGTAGTGGCCGAACCGGTAGAAGTAGTACTCTGCTC	1620						
QY	1621	agcctgtcaaaccaatggctcccaacgcgcatgaaacttttctgtgatatcttag	1680						
Db	1621	AGCCTGTCAACCAATGGTCTCCAAACGCGAGATCAAACTTTATTCTGTGATATCTTAG	1680						
QY	1681	ataacgatttaacgtatgcagacattacatcctttgagccattatttaacaaactccta	1740						
Db	1681	ATAACGATTTAAAGTATGCAGACATTTACATCTCTTTTGAGCCATTTATTAACAACTCTCA	1740						
QY	1741	aggatcctgatgcagagagagctgtaacagtaccatcaaggaagcacctgtacaagtac	1800						
Db	1741	AGGATCTCTGATCCAGAGAGGCTGTAAACAGTACCATCAAGAGGAGCACCTGTACAAAGTAC	1800						
QY	1801	cagtgagtagggccgcgaagaagtgccaacgggaagaattgtagcaactccaagagg	1860						
Db	1801	CAGTGGCAGTAGGGCCGCGCAAGAAAGTGCCAACGGAAGAAATTTGATGCAACTCCAAGAGG	1860						
QY	1861	agcatttcaatttagaagaactgcagaagctccaagagaaggaattagattatagaag	1920						
Db	1861	AGCATTTCAATTTAGAAGGAACCTGCAGAAAGCTCCAGGAAGGAGAAATTAGTATTAGAAG	1920						
QY	1921	ggaaggaagaacacgcgaaggaagccagaagagagagccaaacagaaggaagagtg	1980						
Db	1921	GAGAAGGAGAACCAACGGAAGAGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGTGC	1980						
QY	1981	cagaagaagaatttagaggcaactccagaggaagcatttcgaatttagaagaaccaacaggag	2040						
Db	1981	CAGAAGAAGAAATTAGAGGCAACTCCAGAGGACGATTTTCGAATTTAGAAAGAACCAACAGAGG	2040						
QY	2041	aagaagtaagaagaacccgtagaggcggaagaactccaagagaaggaagtggaagagg	2100						
Db	2041	AAGAAGTAAAGAAACCGTAGAGGGCGAAGAACTGCAGAAAGGAGAGAGTGGAGAGG	2100						
QY	2101	tacctcagaagttagaagaagtgaagaggtacctcagaagttagaagaagtgaagagg	2160						
Db	2101	TACCTCCAGAGTAGAAGAGTGGAAAGAGTACCTCTCCAGAGTAGAAGAGTGGAGAGG	2160						

Qy	2761	cagaatatagtcataattcaacaaaggaaagaccagctggtttgagaaacgagagt	2820
Db	2761	CAGAATATAGTCAAAATTTCAACAAAAGAAAGACCCAGCTGGTTTGAGAAAACGAGAGT	2820
Qy	2821	ggagcagcagagaatggaaaaatggtttaagcagagaagtcaaatccccaaattgattcag	2880
Db	2821	GGAGCCAGCAGAGAAATGGAAAAATGGTTTAAAGCAGAACTCAAAATCCCAAAATGGATTCAAC	2880
Qy	2881	acttgaaaaatgatgaacgacactcattccaattatttaaaattcttctgtaagagata	2940
Db	2881	ACTTGAAGAAATGATGAACGACACTCATTCCATTTTATTAATCTTGTTGAAAGATA	2940
Qy	2941	tgtcacatttgaacaaagaaaccaaagaatggttaatgaatcactcgaaagaaacg	3000
Db	2941	TGTCACAATTTGAAAAACAAGAAACCAAGAAATGGTTTAAATGAATCACTGCAAAAACAACG	3000
Qy	3001	aacggggttatggttcgaatcatttgaagtattgacccacatacaaatattataatgtgg	3060
Db	3001	ACGGGGTTATGGTTCTGAATCAATTTGAAGTTATGACCACATCAAAATTTATTAAATGTGG	3060
Qy	3061	ctaagagtcgagaatggtaccgtgccaatcctaataataatagagaaagagaaactca	3120
Db	3061	CYAAAGAGTCGAGAAATGGTACCGTGCCCAATCTTAATATAATAAGAGAAAGAACTCA	3120
Qy	3121	tgaatggtttctctaaagaaacgaataatttaggacaaagaaatggaaaaatggact	3180
Db	3121	TGAAATGGTTTCTCTCTTAAAGAAAGAAAGCAATATTTAGGACAAAGAAATGGAAAAATGGACT	3180
Qy	3181	cattgaaaaaagttaaaattttttgttcaattcaattcaattgtgtacaacattttctggaaaa	3240
Db	3181	CATTGAAAAAAGTTAAATTTTGTGTTCAATTCATTTGTTCAACATTTTCTTGAAAAA	3240
Qy	3241	cgcttaaccaagagaatggaatcaatttgttaatgaaataaaagtgttgaattatagaa	3300
Db	3241	CGCTTAACCAAGGAATGGAATCAATTTGTTAATGAATAAAAGCTTTGAATTATAGAA	3300
Qy	3301	aaaagacagattattctctataaataataatttc 3337	
Db	3301	AAAAGAACAGATTATTCTCTTATAAATAAATAATTC 3337	
RESULT 3			
ID	E10125	standard; RNA; UNC; 2069 BP.	
AC	E10125;		
SV	E10125.1		
XX			
DT	08-OCT-1997 (Rel. 52, Created)		
DT	02-SEP-2000 (Rel. 65, Last updated, Version 2)		
XX			
DE	DNA encoding an immunogenicity protein.		
XX			
KW	JP 1995284392-A/1.		
XX			
OS	unidentified		
OC	unclassified.		
XX			
RN	[1]		
RP	1-2069		
RA	Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;		
RT	"GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZON IMMUNOGENIC PROTEIN AND GENE		
RT	RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZON";		
RL	Patent number JP1995284392-A/1, 31-OCT-1995.		
RL	DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST.THE.		
XX			
CC	OS Leucocytozon caulleryi		
CC	PN JP 1995284392-A/1		
CC	PD 31-OCT-1995		
CC	PF 19-APR-1994 JP 1994080643		
CC	PI DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO		
CC	PC C12N15/09,A61K39/015,C12P21/02;		

CC	CC	strandedness: Double;	
CC	CC	topology: Linear;	
CC	PH	Key	Location/Qualifiers
CC	PH	source	1. .2069
CC	FT		/organism="Leucocytozon caulleryi"
CC	FT		/strain="shizuoka"
CC	FT	misc_feature	1. .25
CC	FT		/note="EcoRI adaptor"
CC	FT	misc_feature	2045. .2069
CC	FT		/note="EcoRI adaptor"
CC	FT	misc_feature	25. .2046
CC	FT		/product="immunogenicity protein"
XX			
FH	Key	Location/Qualifiers	
FT	source	1. .2069	
FT		/db_xref="taxon:32644"	
FT		/organism="unidentified"	
XX			
SQ	Sequence	2069 BP; 1216 A; 167 C; 443 G; 243 T; 0 other;	
Query Match			
Best Local Similarity 10.3%; Score 342.8; DB 23; Length 2069;			
Matches 874; Conservative 0; Mismatches 807; Indels 7; Gaps 4;			
Qy	955	ctaaacatgttacgggacagaattgccaatgaaacctgtaccattaccagtggcagaag	1014
Db	383	CAACACATTATGAAGAGAAATAAATTTATAGTAGATTATTTACCAGAAGTAATGTAG	442
Qy	1015	agccgcgcgaagtaccagcgggaagaatttagatgccactccagaggatgccttcgcatag	1074
Db	443	AAGAATCAGATGAACACAGACATATAACATATGAAATAGATAATGACATACAGAAGAGC	502
Qy	1075	atgttacagaattcccgagggaagtagaattagtaattagatgaaggaggaactgaagag	1134
Db	503	ATGAAAAAGTAAACATGAAAGAGAAAAAGAAAGTAACACATGAAAGAAATAGAAAAAG	562
Qy	1135	aatcaacggaagtgggacgaagaccaggaagaccacgaagaattagatgccactccag	1194
Db	563	AGAGCATGAAGAAATTAATACATGAAGAGAGAAAAAGAAAGTAACACATGAAGAAATAG	622
Qy	1195	aggatggatttc--gcattagacgaactgcagaagagagaacagagaacacgta--gag	1251
Db	623	AAAAAGACAGCATGAAGAAAGTAAATACATGAAGAGAAAAAGAAAGTAACACATGAAG	682
Qy	1252	gagagaaacagagaagctgcagaagggagaagtatcagaagaacactccagagggaag	1311
Db	683	AAAAAGAAAAAGAGCATGAAGAAAGTAAATACATGAAGAGAAAAAGAAAGTAACAC	742
Qy	1312	aagagttagagcaactccagaggatgatttcgcattagatggaactacattagaagaaa	1371
Db	743	ATGAAGAAATAGAAAAAGAGAGCATGAAGAAAGTAAATACATGAAGAGAAAAAGAAAG	802
Qy	1372	ccgaagaaactgcagaagggagaagaaacccgttagaggggagaagaaacccgtaggggag	1431
Db	803	TAACACATGAAGAAATAGAAAAAGAGAGCATGAAGAAAGTAAATACATGAGGAGAAAAAG	862
Qy	1432	aaacgttagagggagaagctgcagaagggagaagagaagatttagaggcaactccagagg	1491
Db	863	AAGAAGTAACACATGAAGAAATAGAAAAAGAGAGAGCATGAAGAAAGTAAATACATGAAG	922
Qy	1492	atgacttccaatttagaagaaccatcaggaaggagaagggaaggagaaggaaggaagggg	1551
Db	923	AAAAAGAAAGTAAACACATGAAGAAATAGAAAAAGAGAGCATGAAGAAAGTAAATACATG	982
Qy	1552	aaggagaaggagaagcgttagtagcagtgccagtagtgccggaacccggtagaagtagtga	1611
Db	983	AAGAAGAAAAAGAAAGTAACACATGAAGAAAAAGAAAAAGAAAGAGCATGAAGAAAGTAA	1042
Qy	1612	ctcctgctcagcctgtcacaaccaatggtcgctccacgcgcagatgaaactttattcgtg	1671

Db 1043 TACATGAAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGAGTAAACATG 1102
QY 1672 atctttagataacgatttaacgtatgcagacattacatctttagccattattaaac 1731
Db 1103 AAGAAGAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGTAAACATGAAGAAG 1162
QY 1732 aaatctcaaggatcctgtatgcggagaggtcttaacagttaccatcaaggaagccctg 1791
Db 1163 AAAAAGAGAGAGTAAACATGAAGAGAGAGAGAGAGAGTAAACATGAAGAGAGAGAG 1222
QY 1792 tacaagtaccagtgccagtagtggccgcgcaagaagtgcacaagtgccaaagattgatcaac 1851
Db 1223 AAGAGTAAACATGAAGAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGAG 1282
QY 1852 tccaaggagcagatttcgaattagaaggaaactgcagaagctccagaggaaggagaattag 1911
Db 1283 TAACACATGAAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGAGTAAACATG 1342
QY 1912 tattagaaggagagagagaaacacaggaagagagccagagagagagccacagag 1971
Db 1343 AAGAAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGTAAACATGAAGAGAGAG 1402
QY 1972 gagaagtgcacaaaggaagattagagcgaactccagagagcagatttcgaattagaag 2028
Db 1403 AAAAGTAAACATGAAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGAGAG 1462
QY 2029 aaccaacagagagagtagtaagaacacgttagagggcgagaaactgcagaagagagag 2088
Db 1463 AAAAGTAAACATGAAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGAGTAAAC 1522
QY 2089 aagtggagaggttacctgcagaagtagtaagaagtggaagaggttacctgcagaagtagaag 2148
Db 1523 ATGAAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGAGTAAACATGAAGAGAGAG 1582
QY 2149 aagtggagaggttacagaagaagtagtaagaagtagtaccgcagagtagtaagaagtagaag 2208
Db 1583 AAGAAG 1642
QY 2209 aggtaccagaagaagtggagaggttaccagaagaagtggagaggttaccagaagaagtgg 2268
Db 1643 AAGAAGAGAGAGAGTGAAG 1702
QY 2269 aagaggtaccagaagaagtggagagagtagtaagaagtagaagaagtagaggttaccagcgg 2328
Db 1703 AAAATAG 1762
QY 2329 tagtagaagtagaagtaccagcgttagtagaagaagaggtggcagagaagtagtagaagaag 2388
Db 1763 AAGAAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAG 1822
QY 2389 aagaagaaggaggaaccagtagtagaagaagaatgtattacaattagtaataaccatcgg 2448
Db 1823 ATGTAAACATATGAAGAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGAGTAAAC 1882
QY 2449 aagaagatatataattagacaacaaagaaagaaagaaagaaagaaagaaagaaagaaag 2508
Db 1883 ATGAAGAGAGAGAGAGAGTAAACATGAAGAGAGAGAGAGAGTAAACATATGAAGAGAGAG 1942
QY 2509 tcattcgactgcactaccagacgttccaaaggaatttatggagagaagaagaagaactg 2568
Db 1943 AAGAAAAGTAAACATGAAG 2001
QY 2569 cagtgatccattgaaaccagagaagattttgcaaggaagaaattaccatctacagaatggc 2628
Db 2002 AAGTAT 2061
QY 2629 tcacattc 2636
Db 2062 TCGAATTC 2069

RESULT 4
E10126

ID E10126 standard; DNA; UNC; 3399 BP.
XX E10126;
XX E10126;
SV E10126.1
XX 08-OCT-1997 (Rel. 52, Created)
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX DNA encoding an immunogenicity protein of Leucocytozon caulleryi fused to
DE maltose-binding protein.
XX JP 1995284392-A/2.
XX unidentified
OS unclassified.
XX [1]
RN 1-3399
RP Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;
RA "GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZON IMMUNOGENIC PROTEIN AND GEN
RT "RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZON";
RL Patent number JP1995284392-A/2, 31-OCT-1995.
XX DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST.THE.
CC OS None
CC Artificial sequences.
CC PN JP 1995284392-A/2
CC PD 31-OCT-1995
CC PF 19-APR-1994 JP 1994080643
CC PI DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO
CC PC C12N15/09,A61K39/015,C12P21/02;
CC CC strandedness: Double;
CC topology: Linear;
CC FH Key Location/Qualifiers
CC FH source 1..3399 /organism="Artificial sequences"
CC FT CDS 1..3399 /product="fusion protein of maltose-binding
CC FT protein and an immunogenicity protein"
CC FT misc_feature 1..1149 /note="maltose-binding protein"
CC FT misc_feature 1150..1174 /note="EcoRI adaptor"
CC FT misc_feature 1174..3195 /note="immunogenicity protein"
CC FT misc_feature 3194..3218 /note="EcoRI adaptor"
CC FT misc_feature 3219..3399 /note="sequence derived from pMAL-c vector"
XX Key Location/Qualifiers
FH source 1..3399 /db_xref="taxon:32644"
FT /organism="unidentified"
XX Sequence 3399 BP; 1577 A; 509 C; 797 G; 516 T; 0 other;

Query Match 10.3%; Score 342.8; DB 23; Length 3399;
Best Local Similarity 51.8%; Pred. No. 2.6e-40;
Matches 874; Conservative 0; Mismatches 807; Indels 7; Gaps 4;
QY 955 cttaaacatgttaccgggacagaagtgcacaaatgaaacctgtaccattaccagtggcagaag 1014
Db 1532 CAACACATTATGAAGAGAGAGAGTAAATTTATATAGTAGATTATTACCAAGAGTAAATGTAG 1591
QY 1015 agcccgcgcaagtaccagcggaagaattagatgccactccagagagtagcttcgcattag 1074
Db 1592 AGAATCAGATGAACAGAGACATATACATATGAATATAGATATGACATACAGAGAGGC 1651

SplicePredictor (Volker Brendel, Stanford University, <http://gremlini.zool.iastate.edu/cgi-bin/sp.cgi>).
Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.
The 5' clone is T89 and the 3' clone is T7B9.
Location/Qualifiers
1. .111554
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="F1D9"
/clone_lib="IGF BAC"
complement(join(4747. .5195,5297. .5754,5827. .6284, 6457. .6803,6864. .7079,7216. .7286,7326. .7810,7914. .8253, 8311. .8737,9083. .9347))
/note="gene_id:F1D9.1"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="helicase-like protein"
join(11589. .11736,11836. .12019,12305. .12497,12774. .12896, 13021. .13080)
/note="gene_id:F1D9.2"
/codon_start=1
/evidence=not_experimental
/product="ATP-dependent CLPB protein-like"
/protein_id="BAA97091.1"
/db_xref="GI:8777573"
/translation="MLEQILHERVIAQDLDESVDVADIRCSKAGISDPNRPITAFMFMGPTVSQVLGASVGVGDEGKSTEVVRRPYSVVQDEIEKPHNPVSILLQLLDGRITDSHSLHILITIRNEDIKAVYEMKQOVVELARKTFKPKFNNRIDEIVTSQPLNSSVSKIVELOVQMRQMKRLEQPNKLEITLVNLKGFDAENFILLDAQPFNKLVRKMENNAHVKN"
complement(join(19018. .19216,19317. .19390,19600. .19746))
/note="contains similarity to unknown protein
gb|AAC97228.1
gene_id:F1D9.3"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97092.1"
/db_xref="GI:8777574"
/translation="MAKERHVVGVGREEDVKRMKLIMPYLINESLHQOLRALFSGYPSTLKITNYWKELEEMKHTKQEFIAFSRSGFSTGAYIPMCAKIINKDFGKHGLAVLPGIKITLFTLVSKSKLGIKDKLIIIVIVIKLVMP"
join(21817. .22004,22075. .22143)
/note="contains similarity to ATP-dependent CLPB protein
gene_id:F1D9.4"
/codon_start=1
/pseudo
/evidence=not_experimental
complement(22489. .23812)
/note="contains similarity to retroelement pol
polyprotein
gene_id:F1D9.5"
/codon_start=1
/pseudo
/evidence=not_experimental
join(24404. .24543,24659. .24879)
/note="contains similarity to ATP-dependent CLPB protein
gene_id:F1D9.6"
/codon_start=1
/pseudo
/evidence=not_experimental
complement(25073. .29379)
/note="gene_id:F1D9.7"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="retroelement pol polyprotein-like"

CDS
join(30097. .30190,30265. .30551,30614. .31510,31758. .32036)
/note="gene_id:F1D9.8"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="ATP-dependent CLPB protein-like"
join(40639. .40982,41779. .42326,42393. .42794,42903. .42938, 43020. .43142,43226. .43333,43425. .43481,43588. .43650, 43733. .43973,44026. .44107,44503. .44674,44813. .44940, 45043. .45198)
/note="gb|AAD25852.1
gene_id:F1D9.9
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
join(46889. .47780,47984. .48332,48695. .48783)
/note="gene_id:F1D9.10
pir||T01475
similar to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
join(51111. .51234,51470. .52398,52474. .52672,52762. .53360, 53433. .53558,53717. .53809,53908. .54010,54110. .54210, 54472. .54531,54627. .54728,54795. .55047,55142. .55226, 55334. .55637,57320. .57389,57601. .58187)
/note="gene_id:F1D9.11
pir||T02599
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
complement(join(58782. .58995,59066. .59150,59231. .59498, 60276. .61922,61970. .62259,62305. .62482))
/note="gene_id:F1D9.12"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="maize transposon MUDR protein-like"
complement(join(63866. .64008,64160. .64226,64346. .64414, 64504. .64620,64743. .64790,64956. .65053,65183. .65264, 65486. .65566,65677. .65718,65854. .65952,66106. .66144))
/note="contains similarity to human diaphanous 1
gene_id:F1D9.13"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01984.1"
/db_xref="GI:9294133"
/translation="MLSKVKIPLPDLMSVLADESVIDVDQVNLIKFCPTKEAEALKGFIGNKETLGRCEOFFLELLKVRVETKLVFKIQFHSQVSGSKLRIMQITILSIGNALNHGTARGSAIGFHLDSLLKLTDTSRNSKMTLMHLYLCKGVLAEPGLLNFPKDMIQLYLAEMQATSKGLEKVVQEFETASSETDCQTSKHFHMLKEFLSVASGEVRSLSLYSTVGSADALALYFGEDPARVPFEQVSTLQNFVRIEVRSHSEENCKQVEFEKKRAGKEAENEKIKKGVYNE"
complement(66524. .66974)
/note="gb|AAD23008.1
gene_id:F1D9.14
similar to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
complement(join(68701. .69561,69697. .70591,70640. .70667, 71526. .72709,72898. .73300,73370. .73533,73687. .74594))
/note="gene_id:F1D9.15"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="retroelement pol polyprotein-like"
complement(75178. .76058)
/note="gb|AAD23008.1
gene_id:F1D9.16
similar to unknown protein"

```

/codon_start=1
/evidence=not_experimental
76515..80547
/notes="emb|CAB75469.1
gene_id:FID9.17"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="copla-type reverse transcriptase-like protein"
complement(join(81294..81521,81620..82285))
/notes="gb|AD23008.1
gene_id:FID9.18
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
complement(join(87685..88025,88101..88185,88231..88862,
88953..89075,89158..89715))
/notes="gene_id:FID9.19"

Query Match          9.9%  Score 330.6;  DB 8;  Length 111554;
Best Local Similarity 64.9%  Pred. No. 1.3e-38;
Matches 489;  Conservative 0;  Mismatches 264;  Indels 0;  Gaps 0;

Qy 1749 gatgcaggaggctgttaacagttaccatcaagaagacacctgtacaagtaccagtggca 1808
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109928 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 109987

Qy 1809 gtaggcccgccgaagaagtgcccaacggaagaattgtgcaactccaaggaggcatttc 1868
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109988 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110047

Qy 1869 gaattagaaggaactgcagaagctccagaggaggaagaaattagattagaagagaagga 1928
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110048 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110107

Qy 1929 gaaccaacggaagaagcccaaggaagagagcccaacagaagaagagagagtgccagaaga 1988
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110108 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110167

Qy 1989 gaattagaggcaactcagaggagcatttcgaattagaagaaccaacaggagaagaagta 2048
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110168 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110227

Qy 2049 gaagaacccgtagaggcgcaagaactgcagaaggaagaagtggaagaggttaccttqca 2108
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110228 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110287

Qy 2109 gaagtagaagaagtgggaaggggtacctgcagaagtagaagaagtgggaaggggtaccagaa 2168
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110288 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110347

Qy 2169 gaagtagaagaaggttaccgcagaaagttagaagaagtggaagaggttaccagaagaagtggaa 2228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110348 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110407

Qy 2229 gaggtaccagaagaagtgggaaggggtaccagaagaactggaagaggttaccagaagaagtg 2288
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110408 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110467

Qy 2289 gaagaagtggaaagtagaagaagttagaggtaccagcggttagtagaagtagaagtaccac 2348
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110468 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110527

Qy 2349 gcggttagtagaagaagaggtgcccagaagaagtagaagaagaagaagaagaggaagaaacca 2408
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110528 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110587

Qy 2409 gttagagaagaagatgtattacaattagtaataccatcggaagaagatatacaattagac 2468
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110588 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 110647
```

```

Qy 2469 aaaccaagaagaacgaattagctctctggaatt 2501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110648 GAATCGTCACCACATAAATTGGGTATGGGCTT 110680

RESULT 6
AB073159
LOCUS
DEFINITION
Arabidopsis thaliana DNA, chromosome 4 centromere region, BAC
Clone:Tp22.
ACCESSION
AB073159
VERSION
AB073159.1 GI:18149197
KEYWORDS
Arabidopsis thaliana (cultivar:Columbia) DNA, clone:Tp22.
SOURCE
Arabidopsis thaliana
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (sites)
Kumekawa,N., Hosouchi,T., Tsuruoka,H. and Kotani,H.
The size and sequence organization of the centromeric region of
arabidopsis thaliana chromosome 4
Unpublished
2 (bases 1 to 81019)
Kotani,H. and Kumekawa,N.
Direct Submission
Submitted (16-OCT-2001) Hirokazu Kotani, Kazusa DNA Research
Institute, Lab. Chromosome Research II; 1532-3 Yana, Kisarazu,
Chiba 292-0812, Japan (E-mail:kotani@kazusa.or.jp,
Tel:81-438-52-3920, Fax:81-438-52-3921)

FEATURES
source
1..81019
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
/map="centromere"
/misc_feature
1..1562
/clone="Tp22"
/notes="overlap to BAC T24C1, please refer for analysis and
annotation"
misc_feature
1563..3867
/notes="similarity to fragments of Athila retrotransposons"
LTR
3868..5534
misc_feature
5535..6021
/notes="similarity to fragments of Athila retrotransposons"
LTR
6022..7645
misc_feature
7646..12455
/notes="similarity to fragments of Athila retrotransposons"
misc_feature
12456..18907
/notes="similarity to fragments of En/Spm transposon"
misc_feature
18908..20082
/notes="similarity to fragments of Athila retrotransposons"
misc_feature
20083..21988
/notes="similarity to fragments of En/Spm transposon"
misc_feature
21989..26189
/notes="similarity to fragments of Athila retrotransposons"
LTR
26190..27270
misc_feature
27271..27838
/notes="similarity to fragments of Athila retrotransposons"
LTR
27839..29036
misc_feature
29037..30847
/notes="similarity to fragments of En/Spm transposon"
LTR
30848..32651
misc_feature
32652..36522
/notes="similarity to fragments of Athila retrotransposons"
LTR
36523..37042
misc_feature
37043..44711
/notes="similarity to fragments of Tat retrotransposons"
misc_feature
44712..51299
/notes="similarity to fragments of Athila retrotransposons"
LTR
51300..53083
LTR
53084..53829
```

```
misc_feature 53828..58528
/notes="similarity to fragments of Tma retrotransposons"
LTR 58529..59847
misc_feature 59848..62072
/notes="similarity to fragments of Athila retrotransposons"
LTR 62073..62855
LTR 65457..65959
misc_feature 65960..70568
/notes="similarity to fragments of Tat retrotransposons"
repeat_region 70569..71234
/rpt_family="(GAA)n"
misc_feature 71235..73234
/notes="similarity to fragments of Athila retrotransposons"
LTR 73235..75775
LTR 75776..77556
CDS join(78124..78212,78321..78572,78639..78807,79136..79851,
79967..80048)
/codon_start=1
/evidence=not_experimental
/product="putative bZIP transcriptional activator-like
protein"
/protein_id="BAB83611.1"
/db_xref="GI:18149198"
/translation="MQLLCDIDGTYLYKAPKKVILLIYRFVLLTVLSKNVSPAIH
PFIQSLSQNLIEFYNILFTIYPTLLHOOIPHLIKHEIKDKFGONLETKSGHK
PTLKLTAHEHRLPALLNPASLLTSATRKRNLTDSFIRRRQOSIFPFLVLSIQIS
SPSRFTINEMASSKGSQNRNLGCKGKQALLPKSPFTGGTFSADFVSPVSIQSA
VQKLGEGNAHHHTSSEFLIEQPSWLDLNEPPTVRKGGHRRSSDSFAYVDVP
VGFDVDTLWDGGRYNNNGFENHVRPKESDYLRSPQVPPEYPSAHLKQKIRPMDSL
PDSCARPNSSCGLESSITRSCSGSLRDTKAYSAADSKKDFINFAKSFERDND
PLAKSATSEADTKRAQQAQFAQSRVRKIQYIAELERNVQLQV"
BASE COUNT 24352 a 16182 c 17018 g 23467 t
ORIGIN

Query Match          9.5%; Score 318.6; DB 8; Length 81019;
Best Local Similarity 67.3%; Pred. No. 7e-37;
Matches 450; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 1822 aagaagtcgaaggaaggaattgatgcacactccaagggagcatttcgaattagaagaa 1881
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70570 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70629

QY 1882 ctgcagaagctccagaaggaagaattagattagagaaggaaggaaggaaggaag 1941
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70630 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70689

QY 1942 aagagccaaggaaggaagcacaaggaaggaagtcgccaaggaagaattagagcaa 2001
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70690 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70749

QY 2002 ctccagaggacgatttcgaattagaagaaccaaaggaaggaagtagaagaacccgtag 2061
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70750 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70809

QY 2062 agggcgaagaaactgcagaaggaagaagtggaagagggtacctgcagaagtagaagaag 2121
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70810 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70869

QY 2122 tgggaaggggtacctgcagaagtagaagaagtggaagagggtaccagaagaagtagaag 2181
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70870 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70929

QY 2182 taccgcagaagtagaagaagtggaagagggtaccagaagaagtggaagagggtaccagaag 2241
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70930 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70989

QY 2242 aagtcgaagagggtaccagaagaagtggaagagggtaccagaagaagtggaagagggtgaag 2301
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70990 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 71049

QY 2302 aagtagaagaagtagaggtaccagcggtagtagaagtagaagtagaccagcggtagtagaag 2361
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```
Db 71050 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 71109
QY 2362 aagaggtgcagagaagtagaagaagaagaagaagaagaccagtagagaagaag 2421
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 71110 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 71169
QY 2422 atgtattacaattagtaataaccatcggaagaagatatcaattagacaacaaagaaag 2481
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 71170 CAGAAGAAGAAGAAGCAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 71229
QY 2482 acgaattag 2490
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 71230 AAGAATTAG 71238

RESULT 7
AL662884/c 144505 bp DNA linear HTG 08-FEB-2002
LOCUS Homo sapiens chromosome 6 clone Xbac-300A18, *** SEQUENCING IN
DEFINITION PROGRESS ***, 9 unordered pieces.
ACCESSION AL662884
VERSION AL662884.4 GI:18643821
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Sims, S.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 10, 2002 this sequence version replaced gi:18151543.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bpg300A18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 140263 bases at least Q40
Consensus quality: 141340 bases at least Q30
Consensus quality: 142357 bases at least Q20
Insert size: 143705; sum-of-contigs
Insert size: 144041; 6.8% error; agarose-fp
Quality coverage: 9.03x in Q20 bases; sum-of-contigs Quality
coverage: 9.45x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 6934 7033: gap of 100 bp
* 7034 28893: contig of 21860 bp in length
* 28894 28993: gap of 100 bp
* 28994 31651: contig of 2658 bp in length
* 31652 31751: gap of 100 bp
* 31752 45643: contig of 13892 bp in length
* 45644 45743: gap of 100 bp
* 45744 61827: contig of 16084 bp in length
* 61828 61927: gap of 100 bp
* 61928 107959: contig of 46032 bp in length
* 107960 108059: gap of 100 bp
```


* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 23438: contig of 23438 bp in length
* 23439 23538: gap of 100 bp
* 23539 24096: contig of 558 bp in length
* 24097 24196: gap of 100 bp
* 24197 24746: contig of 550 bp in length
* 24747 24846: gap of 100 bp
* 24847 25309: contig of 463 bp in length
* 25310 25409: gap of 100 bp
* 25410 26051: contig of 642 bp in length
* 26052 26151: gap of 100 bp
* 26152 27047: contig of 896 bp in length
* 27048 27147: gap of 100 bp
* 27148 27754: contig of 607 bp in length
* 27755 27854: gap of 100 bp
* 27855 28543: contig of 689 bp in length
* 28544 28643: gap of 100 bp
* 28644 29290: contig of 647 bp in length
* 29291 29390: gap of 100 bp
* 29391 30194: contig of 804 bp in length
* 30195 30294: gap of 100 bp
* 30295 30697: contig of 403 bp in length
* 30698 30797: gap of 100 bp
* 30798 31446: contig of 649 bp in length
* 31447 31546: gap of 100 bp
* 31547 32368: contig of 822 bp in length
* 32369 32468: gap of 100 bp
* 32469 33461: contig of 993 bp in length
* 33462 33561: gap of 100 bp
* 33562 34177: contig of 616 bp in length
* 34178 34277: gap of 100 bp
* 34278 35035: contig of 758 bp in length
* 35036 35135: gap of 100 bp
* 35136 35892: contig of 757 bp in length
* 35893 35992: gap of 100 bp
* 35993 36855: contig of 863 bp in length
* 36856 36955: gap of 100 bp
* 36956 37837: contig of 882 bp in length
* 37838 37937: gap of 100 bp
* 37938 38976: contig of 1039 bp in length
* 38977 39076: gap of 100 bp
* 39077 39933: contig of 857 bp in length
* 39934 40033: gap of 100 bp
* 40034 41139: contig of 1106 bp in length
* 41140 41239: gap of 100 bp
* 41240 42285: contig of 1046 bp in length
* 42286 42385: gap of 100 bp
* 42386 43508: contig of 1123 bp in length
* 43509 43608: gap of 100 bp
* 43609 44697: contig of 1089 bp in length
* 44698 44797: gap of 100 bp
* 44798 46002: contig of 1205 bp in length
* 46003 46102: gap of 100 bp
* 46103 47134: contig of 1032 bp in length
* 47135 47234: gap of 100 bp
* 47235 48609: contig of 1375 bp in length
* 48610 48709: gap of 100 bp
* 48710 50120: contig of 1411 bp in length
* 50121 50220: gap of 100 bp
* 50221 51074: contig of 854 bp in length
* 51075 51174: gap of 100 bp
* 51175 52266: contig of 1092 bp in length
* 52267 52366: gap of 100 bp
* 52367 53467: contig of 1101 bp in length
* 53468 53567: gap of 100 bp
* 53568 54859: contig of 1292 bp in length
* 54860 54959: gap of 100 bp
* 54960 56493: contig of 1534 bp in length
* 56494 56593: gap of 100 bp
* 56594 58097: contig of 1504 bp in length

* 58098 58197: gap of 100 bp
* 58198 60019: contig of 1822 bp in length
* 60020 60119: gap of 100 bp
* 60120 61572: contig of 1453 bp in length
* 61573 61672: gap of 100 bp
* 61673 63511: contig of 1839 bp in length
* 63512 63611: gap of 100 bp
* 63612 64527: contig of 916 bp in length
* 64528 64627: gap of 100 bp
* 64628 66684: contig of 2057 bp in length
* 66685 66784: gap of 100 bp
* 66785 68609: contig of 1825 bp in length
* 68610 68709: gap of 100 bp
* 68710 70596: contig of 1887 bp in length
* 70597 70696: gap of 100 bp
* 70697 72728: contig of 2032 bp in length
* 72729 72828: gap of 100 bp
* 72829 76282: contig of 3454 bp in length
* 76283 76382: gap of 100 bp
* 76383 78231: contig of 1849 bp in length
* 78232 78331: gap of 100 bp
* 78332 84930: contig of 6599 bp in length
* 84931 85030: gap of 100 bp
* 85031 90279: contig of 5249 bp in length
* 90280 90379: gap of 100 bp
* 90380 101025: contig of 10646 bp in length
* 101026 101125: gap of 100 bp
* 101126 109913: contig of 8788 bp in length
* 109914 110013: gap of 100 bp
* 110014 120982: contig of 10969 bp in length
* 120983 121082: gap of 100 bp
* 121083 132369: contig of 11287 bp in length
* 132370 132469: gap of 100 bp
* 132470 151295: contig of 18826 bp in length
* 151296 151395: gap of 100 bp
* 151396 177432: contig of 26037 bp in length
* 177433 177532: gap of 100 bp
* 177533 215046: contig of 37514 bp in length.

FEATURES

source

Location/Qualifiers

1. .215046
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-26F2"
/clone_lib="RPC1-11 Human Male BAC"
1. .23438
/note="assembly_fragment"
clone_end:SP6
vector_side:left
23539..24096
/note="assembly_fragment"
24197..24746
/note="assembly_fragment"
24847..25309
/note="assembly_fragment"
25410..26051
/note="assembly_fragment"
26152..27047
/note="assembly_fragment"
27148..27754
/note="assembly_fragment"
27855..28543
/note="assembly_fragment"
28644..29290
/note="assembly_fragment"
29391..30194

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

Query Match

9.1%; Score 304; DB 2; Length 215046;
Best Local Similarity 47.5%; Pred. No. 8.6e-35;
Matches 969; Conservative 0; Mismatches 1065; Indels 6; Gaps 2;

QY 399 gaagacgatgagggaagtcaacgaacgttagatgaggtgaatcattaggaggtttcc 458

Db 164222 GGAGCAGAGGGCAGAGTTCGGGAGCAGGAGGGCGAGGTCCGGAGCAGAGGAGGGGCGAGGT 164281
QY 459 actgaacatattgaaacacagaagatggctcaacagatgatacgcgaacacagaagaagacta 518
Db 164282 CGGGAGCAGGAGGGGCGAGTTCGGGAGCAGGAGGGCGAGGTCCGGAGCAGGAGGGGCA 164341
QY 519 cctgggtgatatgaagagagaaagaagctgycgatataatggaacagggggaagctgggt 578
Db 164342 GGTCCGGAGCAGGAGGGGCGAGTTCGGGAGCAGGAGGGGCGAGGTCCGGAGCAGGAGGG 164401
QY 579 gatttgaacaggggaagaacactgcgatttggaagcaggggaagaacactgcgatttg 638
Db 164402 GCAGGTCCGGGAGCAGGAGGGGCGAGGTTCGGGAGCAGGAGGGGCGAGGTTCGGGAGCAGGA 164461
QY 639 gaagcaggggaagaagctgggtgatttggaaagcaggggaagaacactggcatttggaaagca 698
Db 164462 GGGGCAGGTCCGGGAGCAGGAGGGGCGAGTTCGGGAGCAGGAGGGGCGAGTTCGGGAGCA 164521
QY 699 ggggaagaactggagatgcggaactgaagaaggagcgaactggagatgcggaactgaa 758
Db 164522 GGAGGGCAGGTTCGGGAGCAGGAGGGGCGAGTTCGGGAGCAGGAGGGGCGAGTTCGGGGA 164581
QY 759 aatggaacactgtgatgtagacacagaagatattcaactgtatgagcagaaaaagta 818
Db 164582 GCAGGAGGACAGATGGGGGAGCAGGAGGAGCAGATGCAGAAACAGGAGGAGCAGATGGG 164641
QY 819 catgttctctcaagaataatgtacaacctggccgatagtaattgatgcctcttggaaagt 878
Db 164642 ---GGAGCAGGAGGACAGATGCGGAAGCAGGAGGAGCAGATGGGGGAGCAGGAGCA 164698
QY 879 attttgataaagatataatttttgatcatattaaagatttcagagcactattcgaaacaa 938
Db 164699 GGTTCAGAACAGGAGGAGCAGGTGCAAAAGCAGGAGGAGCAGATGCGGAAGCAGGAGGA 164758
QY 939 atttgccggtactctaaacatttacgggacagaatttgcattgcaactgtaccac 998
Db 164759 GCAGATCGGGAGCAGGAGGAGCAGATGTCGGGAGCAGGAGGAGCAGATGCTGAAGCAGAA 164818
QY 999 ttaccagtggcagaagaccgcgcgaagtaccagcggaagaattagatgcactccacag 1058
Db 164819 GGAGCAGAGGAGCAGGAGGAGCAGCGGGGAGCAGGAGGAGCAGATGCGGAGCAGGA 164878
QY 1059 gatgaacttcgcatagatgtacagaatctcccgaggaagtagaattagattagatgaa 1118
Db 164879 GGAGCAGATCGGGAGCAGGAGGAGCAGATGTCGGGAGCAGGAGGAGCAGATGTCGGGAGCA 164938
QY 1119 gagcgaactgaagaataacacggaagtgggaccaacggaagaagaccacccaagaa 1178
Db 164939 GGAGGACAGATGGGGAAGCAGGAGGAGCAACAGATGTCGGGAGCAGGAGGAGCAATGTCGGGA 164998
QY 1179 ttagatgccactccagaggatgatttcgcatttagacgaactgcagaagagagaacaga 1238
Db 164999 GCAGAGGACAGATGTCGAAGCAGGAGGAGCAAAATCGGGGAGCAGGAGGAGCAGATGCA 165058
QY 1239 agaaacgtagagggagagaagaacacagaagaagctgcagaagagagaagtatacagaagaact 1298
Db 165059 GAAGCAGGAGGAGCAGGTTCGCAAGCAGGAGGAGCAGATGTCGGAAGCAGGAGGAGCAGAT 165118
QY 1299 ccagaagagagaagaagtttagaggaactccagagagatatttcatttagatgaact 1358
Db 165119 CCGGAACAGGAGGAGCAGATGCGGAAGCAGGAGGAGCAAAATGGGGGAGCAGGAGGAGCA 165178
QY 1359 acattagaagaacccgaagaactgcagaagggagaagaacccgttagagggaagaagaacc 1418
Db 165179 AATGGGGAGCAGGAGGAGCAGATGTCGGGAGCAGGAGGAGCAAAATGGGGAGCAGGAAGGA 165238
QY 1419 gtagagggagaagaacccgttagagggagaagaagctgcagaagaagaagaagagtttag 1478
Db 165239 GCAGATCGGAAGCAGGAGGAGCAGATGCGGGGAGCAGGAAGGAGCAGATGCGGAAGCAGGA 165298
QY 1479 gcaactccagagatgaacttcccaattagaagaaccatcaggagaagaggaagggaagga 1538

Db 165299 GGAGCAGATCGGGAACGACGAGGAGCAGATGCGGAAGCAGGAGGAGCAGATGCGGGGAGCA 165358
QY 1539 gaaggagaaggggaagagagaagaagcgtttagtagcagtgccagtagtgggccgaacccg 1598
Db 165359 GGAGGAGCAGGTTCGGGAGCAGGAGGAGCAGATGCGGGGAGCAGGAGCAGATGCGGAA 165418
QY 1599 gtagaagtagtgactcctcctcagcctgtctaaacccaatagtcgtcctccaacggcagatgaa 1658
Db 165419 GCAGGAGAGACAGATGCTGAAGCAGAGGAGCAGATGCGGGGAGCAGGAGGAGCAGATGGG 165478
QY 1659 acttattcctgtgatattcttagataaacgatttaactgtatgcacacattacatcctttgag 1718
Db 165479 GGAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGCAGATGCAGAACGAGGAGGAGCAAT 165538
QY 1719 ccattatttaacaaactcctcaaggatcctgtatgcaggagaggtctgtaacagtaccatca 1778
Db 165539 GGGGGAGCAGAGGAGCAGATGCGGAAGCAGGAGGAGCAAAATGGGGAGCAGGAGGAGCA 165598
QY 1779 aaggaaagaccctgtacaagtaccagtggcagtaggagcccgcaagaagtgcacaacggaa 1838
Db 165599 AATGGGGAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGCAGAGGTGCAAAAGCAGGAGGA 165658
QY 1839 gaattgatcaactccaagaggacgatttcgaatttagaaggaaactgcagaagctccagag 1898
Db 165659 GCAGATCGGAACGACGAGGAGCAGATGGAAGCAGGAGGAGCAGGTGCGGGAGCAAGAGGA 165718
QY 1899 gaaggagaattagatttagaaggagaagagaaccacccaagaaagacgacgaagaagga 1958
Db 165719 GCAGTGGGGAGCAGGAGGAGCAGATGAGGGAGCAGGAGGAGCAAGATGCGGGGAGCAGGA 165778
QY 1959 gagccaaacgaagagaagtgccagaagaagaatttagaaggcaactccagaaggacgatttc 2018
Db 165779 G---GAGCAGATGCAGAACGAGGAGGAGCAGATGCGGGGAGAAAGGAGGAGCAGATGGGGGA 165835
QY 2019 gaattagaagaaccaacaggaagaagtagtagaagaaccgttagagggcgaagaactgca 2078
Db 165836 GCAGGAGGAGCAGATCGGNAAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGCAGATGGG 165895
QY 2079 gaaggagaagaagtggaaaggttacctgcagaagtagaagaagtggaaaggttacctgca 2138
Db 165896 GGAGCAGGAGGAGCAGATGGGGAGCAGGAGGAGGAGCAGATGCGGAAGCAGGAGGAGCAGGT 165955
QY 2139 gaagtagaagaagtggaaaggttaccagaagaagaatagaagaaggttaccgcagaagtagaa 2198
Db 165956 GGGGAGCAGGAGGAGCAGATGCTGAAGCAGGAGGAGCAGATCGGAAGCAGGAGGAGCA 166015
QY 2199 gaagtgaagaggtaccagaagaagtggaaaggttaccagaagaagtggaaaggttaccac 2258
Db 166016 GATCGGAACGAGGAGGAGCAGGTGGGGGAGCAGGAGGAGCAGATGCGGGAGCAGGAGGA 166075
QY 2259 gaagaagtgaagaggttaccagaagaagtggaaagaagtggaaagaattagaagaagtagag 2318
Db 166076 GCAGATGTGGAAGCAGGAGGAGCAGATGGGGGAGCAGGAGGAGCAGGAGGAGCAGATGAG 166135
QY 2319 gtaccagcggtagtagaagtagaagttaccagcgttagtagaagaagaaggtgcacagaagaa 2378
Db 166136 GAAGCAGGAGGAGCAGATGGGGGAGCAGGAGGAGCAGATGGGGAAGCAGGAGGAGCAGAT 166195
QY 2379 gtagaagaagaagaagaggaagaccagtagagaagaagaatgtattaccattagta 2438
Db 166196 GGGGGAGCAGGAGCAGCAGATGCGGGGAGCAGGAGGAGCAGATGCGGGGAGCAGGAGGAGCA 166255

RESULT 10

AC044842

LOCUS

DEFINITION

AC044842

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC044842 Homo sapiens chromosome 2 clone RP11-178E20 map 2, WORKING DRAFT
SEQUENCE, 23 unordered pieces.

AC044842

AC044842.2 GI:8705156

HTG; HTGS_PHASE1; HTGS_DRAFT.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 172307)
 Birren B., Linton L., Nusbaum C. and Lander E.
 Homo sapiens chromosome 2, clone RP11-178E20
 Unpublished
 2 (bases 1 to 172307)
 Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N., Anderson S., Baldwin J., Barna N., Bastien V., Beda F., Boguslavsky L., Boukhgalter B., Brown A., Burkett G., Campopiano A., Castle A., Choepel Y., Colangelo M., Collins S., Collymore A., Cooke P., Dearellano K., Dewar K., Diaz J.S., Dodge S., Domino M., Doyle M., Ferreira P., FitzHugh W., Gage D., Galagan J., Gardyna S., Ginde S., Goyette M., Graham L., Grand-Pierre N., Grant G., Hagos B., Heaford A., Horton L., Howland J.C., Iliev I., Johnson R., Jones C., Kann L., Karatas A., Klein J., LaRocque K., Lamazares R., Landers T., Lehoczeky J., Levine R., Lieu C., Liu G., Locke K., Macdonald P., Marquis N., McCarthy M., McEwan P., McGurk A., McKernan K., McPheeters R., Meldrim J., Meneus L., Mihova T., Miranda C., Mlenga V., Morrow J., Murphy T., Naylor J., Norman C.H., O'Connor T., O'Donnell P., O'Neil D., Oliver T.M., Oliver J., Peterson K., Pierre N., Pisani C., Pollara V., Raymond C., Riley R., Rogov P., Rothman D., Roy A., Santos R., Schauer S., Severy P., Spencer B., Stange-Thomann N., Stojanovic N., Subramanian A., Talamas J., Testaye S., Theodore J., Tirrell A., Travers M., Trigilio J., Vassiliev H., Viel R., Vo A., Wilson B., Wu X., Wyman D., Ye W.J., Young G., Zainoun J., Zimmer A. and Zody M.
 Direct Submission
 Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 25, 2000 this sequence version replaced gi:7543808.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L9002
 Center clone name: 178_E20

----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 156589 bases at least Q40
 Consensus quality: 16477 bases at least Q30
 Consensus quality: 167863 bases at least Q20
 Insert size: 171000; agarose-fp
 Insert size: 170107; sum-of-contigs
 Quality coverage: 3.7 in Q20 bases; agarose-fp
 Quality coverage: 3.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 19: contig of 19 bp in length
 * 20 119: gap of 100 bp
 * 120 1921: contig of 1802 bp in length
 * 1922 2021: gap of 100 bp
 * 2022 3689: contig of 1668 bp in length
 * 3690 3789: gap of 100 bp
 * 3790 6531: contig of 2742 bp in length
 * 6532 6631: gap of 100 bp
 * 6632 8134: contig of 1503 bp in length
 * 8135 8234: gap of 100 bp

* 8235 11919: contig of 3685 bp in length
 * 11920 12019: gap of 100 bp
 * 12020 15244: contig of 3225 bp in length
 * 15245 15344: gap of 100 bp
 * 15345 17344: contig of 2000 bp in length
 * 17345 17444: gap of 100 bp
 * 17445 21830: contig of 4386 bp in length
 * 21831 21930: gap of 100 bp
 * 21931 25081: contig of 3151 bp in length
 * 25082 25181: gap of 100 bp
 * 25182 29736: contig of 4555 bp in length
 * 29737 29836: gap of 100 bp
 * 29837 36853: contig of 7017 bp in length
 * 36854 36953: gap of 100 bp
 * 36954 41270: contig of 4317 bp in length
 * 41271 41370: gap of 100 bp
 * 41371 46765: contig of 5395 bp in length
 * 46766 46865: gap of 100 bp
 * 46866 57601: contig of 10736 bp in length
 * 57602 57701: gap of 100 bp
 * 57702 67879: contig of 10178 bp in length
 * 67880 67979: gap of 100 bp
 * 67980 78235: contig of 10256 bp in length
 * 78236 78335: gap of 100 bp
 * 78336 89176: contig of 10841 bp in length
 * 89177 89276: gap of 100 bp
 * 89277 106081: contig of 16805 bp in length
 * 106082 106181: gap of 100 bp
 * 106182 121387: contig of 15206 bp in length
 * 121388 121487: gap of 100 bp
 * 121488 137207: contig of 15720 bp in length
 * 137208 137307: gap of 100 bp
 * 137308 153745: contig of 16438 bp in length
 * 153746 153845: gap of 100 bp
 * 153846 172307: contig of 18462 bp in length.

----- Location/Qualifiers
 1..172307

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-178E20"
 /clone_lib="RPC1-11 Human Male BAC"

1..19
 /note="assembly_fragment
 vector_side:left"

misc_feature

120..1921
 /note="assembly_fragment"

misc_feature

2022..3689
 /note="assembly_fragment"

misc_feature

3790..6531
 /note="assembly_fragment"

misc_feature

6632..8134
 /note="assembly_fragment"

misc_feature

8235..11919
 /note="assembly_fragment"

misc_feature

12020..15244
 /note="assembly_fragment"

misc_feature

15345..17344
 /note="assembly_fragment"

misc_feature

17445..21830
 /note="assembly_fragment"

misc_feature

21931..25081
 /note="assembly_fragment"

misc_feature

25182..29736
 /note="assembly_fragment"

misc_feature

29837..36853
 /note="assembly_fragment"

misc_feature

36954..41270
 /note="assembly_fragment"

misc_feature

misc_feature	41371..45765	/note="assembly_fragment"		
misc_feature	46866..57601	/note="assembly_fragment"		
misc_feature	57702..67879	/note="assembly_fragment"		
misc_feature	67980..78235	/note="assembly_fragment"		
misc_feature	78336..89176	/note="assembly_fragment"		
misc_feature	89277..106081	/note="assembly_fragment"		
misc_feature	106182..121387	/note="assembly_fragment"		
misc_feature	121488..137207	/note="assembly_fragment"		
misc_feature	137308..153745	/note="assembly_fragment"		
misc_feature	153846..172307	/note="assembly_fragment"		
BASE COUNT	44249 a 40378 c 41592 g 43880 t	2208 others		
ORIGIN				
Query Match 9.08; Score 299; DB 2; Length 172307;				
Best Local Similarity 47.38; Pred. No. 4.5e-34;				
Matches 1032; Conservative 0; Mismatches 1140; Indels 9; Gaps 4;				
Qy	280 atcctcaacgacgcggttttagaattagaagaggaaaaattttgatgagaattcagtgatg	339		
Db	38045 ATAGTTACATGAGCAGTGCAGTAGCAGCGGCAGTAATAGTAGTGTGTAGTAGCAGCAGTTG	38104		
Qy	340 atgaaccttttagatgctaccccgaaagatgactttgcttaacagatttgcaattg	399		
Db	38105 CAGTAGCTGTGGCAGGAGGATGACGACGACGAGTTGCCGTAGCTGTGCAGGAGGAG	38164		
Qy	400 aagacgatgaggaagtcaacgaacgttagtgaggtgaaatcattagagaggtttcca	459		
Db	38165 TAGCAGCAGCAGCAGTTGCCGTAGCTGTGGCAGGAGGAGTAGCAGCAGCAGTTGCC-	38223		
Qy	460 ctgaagatatgaaacagagaagatggctcaacagatgatacggaaacagaagaagactac	519		
Db	38224 --GTAGCTGTGGCAGGAGGATGACGACGACGAGTTGCCGTAGCTGTGCAGGAGGAG	38281		
Qy	520 ctggtgatgaggaagagaagaagctggcgcgataggaaagcaggggaaagaagctggtg	579		
Db	38282 TAGCAGCAGCAGCAGTTGCCGTAGCTGTGGCAGGAGTAGCAGCAGCAGTTGCAGTAG	38341		
Qy	580 atttggaaagcagggaagaaaaactggcgaatttggaaagcaggggaaagaactggcattgg	639		
Db	38342 CTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAG	38401		
Qy	640 aagcaggggaaagaagctggtgatttgaagcaggggaaagaacttggcgatttggaaagcag	699		
Db	38402 CATCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTG	38461		
Qy	700 gggaaagaacttggagatgcgggaaactgaagaagaggaacttggagatgcgggaaactgaaa	759		
Db	38462 TGGCAGGAGTAGCAGCAGCAGTTGCCGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAG	38521		
Qy	760 atgagcaactgttatgttagacagaagatagttcacgtgatggagcagaaaaaagtac	819		
Db	38522 TTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAG	38581		
Qy	820 atgtctctgcctaagaaaaatgacaacctgcgcgatagtaattgatgccctctttggaagta	879		
Db	38582 GAGTAGCAGCATCAGCAGTTGCCGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTG	38641		
Qy	880 ttttggataaagataataatttttgcataataaagatttcgcagccactatttcgaacaaa	939		
Db	38642 AAGTAGCTGTGGCAGGAGGAGTAGCAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAGCAG	38701		
Qy	940 ---ttgtgcgcggttactgcctaacaatgtttacqggacaagaatttgcgaatgaacctgtac	996		

```

Db 39782 TAGCTGTGTTAGAGTAGCAGCATCAGCAGTTGCGAGTAGCTGTGGCAGAGGAGTAGCAG 39841
QY 2074 ctgcagaaggaagaagtggaagaggtacctgcagagatagaagaagtggaagaggtac 2133
Db 39842 CATCACAGTTCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCGAGTCGCTG 39901
QY 2134 ctgcagaaggaagaagtggaagaggtaccagaagaagtagaagaaggtaccgcagaag 2193
Db 39902 TGGCAGGAGGAGTAGCAGCAGCAGCAGTTCAGTAGCTGTGGCAGGAGTAGCAGCATCAG 39961
QY 2194 tagaagaagtggaagaggtaccagaagaagtggaagaggtaccagaagaagtggaagag 2253
Db 39962 TAGTTCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCGAGTAGCTGTGGCAG 40021
QY 2254 taccagaagaagtggaagaggtaccagaagaagtggaagaagtggaagaagtagaagaag 2313
Db 40022 GAGGAGTAGCAGCATCAGCAGTTGCGAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAG 40081
QY 2314 tagaggtaccagcgttagtagaagtagaagtagaccagcgttagtagaagaagaggtgcag 2373
Db 40082 TTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCGAGTAGCTGTGGCAGGAG 40141
QY 2374 aagaagtagaagaagaagaagaagaggaaccagtagagaagaagtagattacaat 2433
Db 40142 TAGCAGCATCAGCAGTTGCGAGCAGCTGTGGCAGGAGGAGTAGCAGCATCAGGAGTTGCAG 40201
QY 2434 tagtaataccatcgggaagaag 2454
Db 40202 CAGCTGTGGCAGGAGGAGTAG 40222

RESULT 11
AC016204
LOCUS
DEFINITION Homo sapiens clone RP11-25C1, *** SEQUENCING IN PROGRESS ***, 5
unordered pieces.
ACCESSION AC016204
VERSION AC016204.6 GI:18644852
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 142277)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome, clone RP11-25C1
Unpublished
2 (bases 1 to 142277)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Trafaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,B., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 11, 2002 this sequence version replaced gi:14209777.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

```

```

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4539
Center clone name: 25_C_1
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence,
* as soon as it is available and the accession number will
* be preserved.
*
* 1 24412: contig of 24412 bp in length
* 24413 24512: gap of 100 bp
* 24513 46948: contig of 22436 bp in length
* 46949 47048: gap of 100 bp
* 47049 103797: contig of 56749 bp in length
* 103798 103897: gap of 100 bp
* 103898 126333: contig of 22436 bp in length
* 126334 126433: gap of 100 bp
* 126434 142277: contig of 15844 bp in length.
*
* Location/Qualifiers
* 1..142277
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /clone="RP11-25C1"
* /clone.lib="RPC1-11 Human Male BAC"
*
BASE COUNT 39176 a 32698 c 33584 g 36253 t 566 others
ORIGIN

Query Match 8.8%; Score 294.6; DB 2; Length 142277;
Best Local Similarity 47.0%; Pred. No. 2e-33;
Matches 932; Conservative 0; Mismatches 1049; Indels 4; Gaps 1;

QY 399 gaagcagctaggaagtcacaaacgcttagatgaggtgaaatcattaggaaggtttcc 458
Db 131759 GGAGCAGGAGGGCGCATGTCGGGAGCAGGAGGGCGGAGGAGGAGGAGGAGGAGT 131818
QY 459 actgaagatatgaaacagaagaatgctcaacagatgatacgaacacagaagaagacta 518
Db 131819 GCGGAGCAGGAGGGCGCATGTCGGGAGCAGGAGGGCGGAGGAGGAGGAGGAGGCA 131878
QY 519 cctgtgtatggaaggaagaagaagctgctgcatatggaagcaggggaagaagctgtg 578
Db 131879 GATGGGGAGCAGGAGGAGCAGATGCAGAAAGCAGGAGGAGCAGATGGGGAGCAGGAGGA 131938
QY 579 gatttggaagcaggggaagaactggcgtatttggaaagcaggggaagaactggcgtattg 638
Db 131939 GCAGATGCGGAAGCAGGAGGAGCAGATGTCGGGAGCAGGAGGAGGAGCAGTGCAGAAAGCAGGA 131998
QY 639 gaagcaggggaagaagctgtgtatttgaagcaggggaagaagaactggcgtatttgaagca 698
Db 131999 GGAGCAGGTGCAAAAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGGAGGAGGAGGAGCA 132058
QY 699 ggggaagaactggagatgcggaactgaagaagggagcaactggagatgcggaactgaa 758
Db 132059 GGAGGAGCAGATGTCGGGAGCAGGAGGAGCAGATGCTGAAGCAGGAGGAGCAGGAGGAGCA 132118
QY 759 aatggagcaactgtgtatgtagacacagaagaatagttcagctgtatggagcagaagaagta 818
Db 132119 GGAGGAGCAGNNNNNNNNNACGGGGGAGCAGGAGGAGGAGCAGATGCGGGAGCAGGAGGAGCA 132178
QY 819 catgttcctgtcagaagaataatgtacaaacctgcgatagtaatgtatgcctctttggaagt 878
Db 132179 GATGCGGGAGCAGGAGGAGCAGATGTCGGGAGCAGGAGGAGGAGGAGGAGGAGGAGCA 132238
QY 879 atttggataaagataataattttatcatattataatttgcagccatttcgaacaa 938
Db 132239 GCAGATGGGGAACAGGAGGAGCAGATGTGGGAGCAGAGGAGCAAAATGTGGGAGCAGAA 132298
QY 939 atttgcgcggtactgctaataacatgttacgggacaagaattgccaatgaacacctgtacca 998

```

Db 132299 GGAGCAGATGGAAGCAGGAGGAGCAAAATGGGGAGCAGGAGGACAGATGCGAGAAGCA 132358
QY 999 ttaccagtgcagaagcccgccgcaagtaccagcggaagaattagatccactccagag 1058
Db 132359 GGAGGACAGGTGCGAAAGCAGGAGGAGCAGATGCGGAAGCAGGAGCAGATGCGGAA 132418
QY 1059 gatgaacttcgattagattgttacagaattctcccgaggaagtagaatttagtattagatgaa 1118
Db 132419 GCAGGAGGACAGATCGGGAAGCAGGAGGAGCAAAATGGGGAGCAGAGAAGGACAAATGGG 132478
QY 1119 gaggcaactgaagaagaatacaacggaagtgggaccacgaagaagaagcccaaccgaaga 1178
Db 132479 GGAGCAGGAGGACAGATCGCGGAGCAGGAGGAGCAAAATGGGGAGCAGAGAAGGAGCAGAT 132538
QY 1179 ttatgttgcactccagagatgatttcgatttagacgaactgcagaaggaagaacaga 1238
Db 132539 CGCGNAGCAGGAGGACACATGGGGAGCAGNAGAAAGGAGCAGATGCGGAAGCAGGAGGAGC 132598
QY 1239 aga-----aacgtagagggagaagaacagaagaagctgcagaaggaaggaagatccagaaga 1294
Db 132599 AGATGCGGAAGCAGGAGGAGCAGATCGGAAGCAGGAGGAGCAGATGGGGAGCAGGAGG 132658
QY 1295 aactccagaaggaagaagagtagtaggcaactccagaggatgatttcgattagatgg 1354
Db 132659 AGCAGGTGCGGGAGCAGGAGGAGCAGATGGGGAGCAGGAGGAGCAGATGCGGAAGCAGG 132718
QY 1355 aactacattagaagaacccaagaactgcagaaggaagaacacccgtagagggagaaga 1414
Db 132719 AGAAGCAGATGCTGAAGCAGAAAGGACAGATGGGGAGCAGGAGGAGCAGATGGGGAGC 132778
QY 1415 aaccgtagagggagaagaaccctagagggagaagaagctgcagaaggaagaagaagtt 1474
Db 132779 AGGAGGAGCAGATGCGGAAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGCAAAATGGGG 132838
QY 1475 agagcaactccagagatgacttccaattagaagaacccaatcaggaaggaagaagggga 1534
Db 132839 AGCAGAAGGAGCAGATGCGGAAGCAGGAGGAGCAAAATGGGGAGCAGGAGGAGCAAAATGG 132898
QY 1535 aggaaggaaggaagggagaaggaaggaagcgttagtcagctgcagagtcaccagttagtgcgca 1594
Db 132899 GGGAGCAGGAGGAGCAGATGCGGAAGCAGGAGCAGGAGGTGCAAAAGCAGGAGGAGCAG 132958
QY 1595 accggtagaagtagtagtactctctcagcctgtcaaacccaatggtcgtcccaacggcaga 1654
Db 132959 TGCGGGAAGCAGGAGGAGCAGATGCGAGCAGGAGCAGGTGCGGGAGCAGAGAGGAGCAGG 133018
QY 1655 tgaactttattcgttatcttagataacgatttaacgtagtcagacattacatcctt 1714
Db 133019 TGGGGAGCAGGAGGAGCAGATGAGGAGCAGGAGGAGAGATGGGGAGCAGGAGGAGC 133078
QY 1715 tgagccattatttaacaaatcctcaggatcctgtcagggagagcgtgttaacagtacc 1774
Db 133079 AGATGCAGAAGCAGGAGGAGCAGATGGGGAGGAAGGAGGAGCAGATGGGGAGCAGGAGG 133138
QY 1775 atcaaggaagcaccctgtacaagtaccagtggcagttagggccgcgaagaagatgcacaac 1834
Db 133139 AGCAGATGCGGAAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGGAGGAGGAGGAGCAGG 133198
QY 1835 ggaagaattgatcaactccaagagacgatttcggaattagaaggaactgcagaagctcc 1894
Db 133199 AGAGCAGATGCGGGAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGGAGCAGATGGGGAGC 133258
QY 1895 agagggaaggaattagattagaaggaagaaggaaggaacccaacggaagaagagccaagaga 1954
Db 133259 AGGAGGAGCAGATGAGGAAGCAGGAGGAGCAGATGGGGAGCAGGAGGAGCAGATGGGA 133318
QY 1955 agagagccaacagaaggaagaagtcgaagaagaattagadggcaactccagagagaca 2014
Db 133319 AGCAGGAGGAGCAGATGGGGAGCAGGAGGAGCAGATGGGGAGGAGGAGGAGGAGCAGATGG 133378
QY 2015 tticgaattagaagaaccaacagagaagaagtagaagaacccgtagagggcgaagaac 2074

Db 133379 GGGAGCAGGAGGAGCAGATGTGGGAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGCAG 133438
QY 2075 tgcagaaggaagaagtgaaagaggtacctgcagaagtagaagaagtggaagaggtacc 2134
Db 133439 TGGGGAGCAGGAGGAGCAGATGAGGAATCAGGAGGAGCAGATGGGGAGCAGGAGGAGC 133498
QY 2135 tgcagaagtagaagaagtgaaagaggtaccagaagaagtagaagaaggtaccctccgcagaagt 2194
Db 133499 AGATGCGGAAGCAGGAGGAGCAGATGGGGAGCAGGAGGAGCAGATGAGGAACGACGAGG 133558
QY 2195 agaagaagtggaaagaggtaccagaagaagtggaagaaggtaccagaagaagtggaagaaggt 2254
Db 133559 ACCAGATGGGGAGCAGGAGGAGCAGATCGGAAGCAGGTGCGGATTCGGAAGG 133618
QY 2255 accagaagaagtggaaagaggtaccagaagaagtggaagaagtggaagaagtagaagaagt 2314
Db 133619 AGGAGAGGCTGTGGGATGAGTATGAGAAGATCCAGGAGGAGGAGAGATCCGGGGGC 133678
QY 2315 agaggtaccagcggttagtagaagtagaagttaccagcggttagtagaagaagagtgccaga 2374
Db 133679 AGGTGAGAGAAGATGGGGGAGAGAAGAGGAGGATGGGAGCAGGAGGAGAGATGTCAGG 133738
QY 2375 aqaag 2379
Db 133739 AGGAG 133743

RESULT 12
AF202180
LOCUS AF202180 16377 bp DNA linear INV 02-APR-2001
DEFINITION Plasmodium falciparum erythrocyte membrane-associated giant protein antigen 332 gene, complete cds.
ACCESSION AF202180 AF327072 AH010491 AF323001 AF323000 AF322999
VERSION AF202180.3 GI:13508498
KEYWORDS malaria parasite P. falciparum.
SOURCE Plasmodium falciparum
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 16377)
AUTHORS Shan,Z.X., Yu,X.B., Li,X.R., Ma,C.L., Fang,J.M., Lu,J.H. and X.J.
TITLE Molecular cloning and structure analysis of the Plasmodium falciparum erythrocyte membrane-associated giant protein Ag332 (Pf332) gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 16377)
AUTHORS Shan,Z.X., Yu,X.B., Li,X.R. and Ma,C.L.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Parasitology, Sun Yat-Sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, China
REFERENCE 3 (bases 1 to 16377)
AUTHORS Shan,Z.X.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-2000) Parasitology, Sun Yat-Sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, China
REMARK Sequence update by submitter
REFERENCE 4 (bases 1 to 16377)
AUTHORS Shan,Z.X.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2001) Parasitology, Sun Yat-Sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, China
REMARK Sequence update by submitter
REFERENCE 5 (bases 1 to 16377)
AUTHORS Shan,Z.X.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-2001) Parasitology, Sun Yat-Sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, China
REMARK Sequence update by submitter
COMMENT On Apr 2, 2001 this sequence version replaced gi:13111670 gi:12232085 gi:13177607 gi:12719373.

[illegible]

REFERENCE
AUTHORS
Rattus.
 1 (bases 1 to 102195)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Banks,T., Barbaris,J.,
 Benton,S., Brieva,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowles,S., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Deigado,O.,
 Dunn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Duggan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulsegad,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newtonson,J., Newtonson,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwuo,S.,
 Ogum,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
 Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 and Gibbs,R.
Direct Submission
 Unpublished
 2 (bases 1 to 102195)
 Worley,K.C.
Direct Submission
 Submitted (18-APR-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 21, 2001 this sequence version replaced gi:13661904.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: TUGG
 Center clone name: CH230-1G12
 ----- Summary Statistics
 Sequencing vector: Plasmid: M77789
 Sequencing vector: M13; L08821
 Chemistry: Dye-primer Bodipy: 65% of reads
 Chemistry: Dye-terminator Big Dye: 35% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 99820 bases at least Q40
 Consensus quality: 116847 bases at least Q30
 Consensus quality: 125383 bases at least Q20
 Estimated insert size: 90042; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 *

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 55 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 3199: contig of 3199 bp in length
 * 3200: gap of unknown length
 * 3300: contig of 2586 bp in length
 * 5886: gap of unknown length
 * 5886: contig of 2677 bp in length
 * 8663: gap of unknown length
 * 8762: contig of 2723 bp in length
 * 11485: contig of 2723 bp in length
 * 11486: gap of unknown length
 * 11585: contig of 2248 bp in length
 * 13933: gap of unknown length
 * 13934: contig of 2583 bp in length
 * 16516: gap of unknown length
 * 16517: contig of 2975 bp in length
 * 19591: gap of unknown length
 * 19592: contig of 2320 bp in length
 * 19692: gap of unknown length
 * 22012: contig of 2432 bp in length
 * 22112: gap of unknown length
 * 24544: contig of 1734 bp in length
 * 24644: gap of unknown length
 * 26378: contig of 2905 bp in length
 * 26478: gap of unknown length
 * 29382: contig of 2661 bp in length
 * 29482: gap of unknown length
 * 32143: contig of 2238 bp in length
 * 32144: gap of unknown length
 * 32443: contig of 2238 bp in length
 * 34481: gap of unknown length
 * 34581: contig of 2338 bp in length
 * 36919: gap of unknown length
 * 37019: contig of 1896 bp in length
 * 38915: gap of unknown length
 * 39015: contig of 1976 bp in length
 * 39016: gap of unknown length
 * 40992: contig of 1823 bp in length
 * 41092: gap of unknown length
 * 42914: contig of 2513 bp in length
 * 43014: gap of unknown length
 * 43015: contig of 1555 bp in length
 * 45227: gap of unknown length
 * 45628: contig of 1197 bp in length
 * 47182: gap of unknown length
 * 47183: contig of 1197 bp in length
 * 48479: gap of unknown length
 * 48580: contig of 2768 bp in length
 * 51347: gap of unknown length
 * 51447: contig of 1384 bp in length
 * 52831: gap of unknown length
 * 52931: contig of 1269 bp in length
 * 54200: gap of unknown length
 * 54201: contig of 2114 bp in length
 * 56414: gap of unknown length
 * 56514: contig of 1274 bp in length
 * 57788: gap of unknown length
 * 57889: contig of 1432 bp in length
 * 59320: gap of unknown length
 * 59321: contig of 1138 bp in length
 * 60558: gap of unknown length
 * 60559: contig of 1897 bp in length
 * 62555: gap of unknown length
 * 62556: contig of 1366 bp in length
 * 64021: gap of unknown length
 * 64022: contig of 1546 bp in length
 * 64122: gap of unknown length
 * 65668: contig of 1837 bp in length
 * 65768: gap of unknown length
 * 67704: contig of 1536 bp in length
 * 69240: gap of unknown length
 * 69241: contig of 1307 bp in length
 * 70647: gap of unknown length

[illegible]

Quality coverage: 6.8 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 30267: contig of 30267 bp in length
* 30268: gap of unknown length
* 30288: contig of 28113 bp in length
* 58401: gap of unknown length
* 58421: contig of 30444 bp in length
* 88864: gap of unknown length
* 88865: contig of 21005 bp in length
* 109889: contig of 21005 bp in length
* 109890: gap of unknown length
* 109910: contig of 15224 bp in length
* 125133: gap of unknown length
* 125134: contig of 14396 bp in length
* 139549: gap of unknown length
* 139550: contig of 13841 bp in length
* 139570: gap of unknown length
* 153410: contig of 9137 bp in length
* 153431: gap of unknown length
* 162567: contig of 9137 bp in length
* 162568: gap of unknown length
* 162587: contig of 8457 bp in length
* 171044: gap of unknown length
* 171064: gap of unknown length
* 171065: contig of 8240 bp in length
* 173305: gap of unknown length
* 179324: contig of 7735 bp in length
* 187059: gap of unknown length
* 187060: contig of 10519 bp in length
* 187080: gap of unknown length
* 197518: gap of unknown length
* 197599: contig of 8348 bp in length
* 205966: gap of unknown length
* 205967: gap of unknown length
* 205987: contig of 6463 bp in length
* 212450: gap of unknown length
* 212469: contig of 5283 bp in length
* 217470: gap of unknown length
* 217752: contig of 6171 bp in length
* 217753: gap of unknown length
* 223943: contig of 6399 bp in length
* 223944: gap of unknown length
* 223963: contig of 6399 bp in length
* 230362: gap of unknown length
* 230363: contig of 5869 bp in length
* 230383: gap of unknown length
* 236251: contig of 4301 bp in length
* 236252: gap of unknown length
* 236272: contig of 2197 bp in length
* 240573: gap of unknown length
* 240593: contig of 1886 bp in length
* 242790: gap of unknown length
* 242810: contig of 1886 bp in length
* 244695: gap of unknown length
* 244715: contig of 2254 bp in length
* 244716: gap of unknown length
* 246970: gap of unknown length
* 246989: contig of 881 bp in length
* 246990: gap of unknown length
* 247870: contig of 303 bp in length
* 247890: gap of unknown length
* 248193: contig of 1909 bp in length
* 248194: gap of unknown length
* 250122: contig of 1225 bp in length
* 250123: gap of unknown length
* 250143: contig of 1372 bp in length
* 251367: gap of unknown length
* 251387: contig of 2281 bp in length
* 251388: gap of unknown length
* 252749: contig of 2056 bp in length
* 252750: gap of unknown length
* 252759: contig of 871 bp in length
* 252770: gap of unknown length
* 253641: contig of 1372 bp in length
* 253660: gap of unknown length
* 255032: contig of 1372 bp in length
* 255033: gap of unknown length
* 255053: contig of 2281 bp in length
* 257333: gap of unknown length
* 257334: contig of 2056 bp in length
* 257354: gap of unknown length
* 259410: contig of 1030 bp in length
* 259429: gap of unknown length
* 260459: contig of 1030 bp in length

* 260460 260479: gap of unknown length
* 260480 260829: contig of 350 bp in length
* 260830 260849: gap of unknown length
* 260850 261661: contig of 812 bp in length
* 261662 261681: gap of unknown length
* 261682 262532: contig of 851 bp in length
* 262533 262552: gap of unknown length
* 262553 263351: contig of 799 bp in length
* 263352 263371: gap of unknown length
* 263372 264217: contig of 846 bp in length
* 264218 264237: gap of unknown length
* 264238 265205: contig of 968 bp in length
* 265206 265225: gap of unknown length
* 265226 266270: contig of 1045 bp in length
* 266271 266290: gap of unknown length
* 266291 266749: contig of 459 bp in length
* 266750 266769: gap of unknown length
* 266770 267718: contig of 949 bp in length
* 267719 267738: gap of unknown length
* 267739 268576: contig of 838 bp in length
* 268577 268596: gap of unknown length
* 268597 270010: contig of 1414 bp in length
* 270011 270030: gap of unknown length
* 270031 271654: contig of 1624 bp in length
* 271655 271674: gap of unknown length
* 271675 272545: contig of 871 bp in length.
* 271675 Location/Qualifiers
1. .272545
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/clone="RP23-33H13"
/sex="male"
1. .30267
/note="assembly_name:Contig108"
30288. .58400
/note="assembly_name:Contig107"
clone_end:SP6
vector_side:left
58421. .88864
/note="assembly_name:Contig106"
88885. .109889
/note="assembly_name:Contig105"
109910. .125133
/note="assembly_name:Contig104"
clone_end:T7
vector_side:right
125154. .139549
/note="assembly_name:Contig103"
139570. .153410
/note="assembly_name:Contig102"
153431. .162567
/note="assembly_name:Contig101"
162588. .171044
/note="assembly_name:Contig100"
171065. .179304
/note="assembly_name:Contig99"
179325. .187059
/note="assembly_name:Contig98"
187080. .197598
/note="assembly_name:Contig97"
197619. .205966
/note="assembly_name:Contig96"
205987. .212449
/note="assembly_name:Contig95"
212470. .217752
/note="assembly_name:Contig94"
217773. .223943
/note="assembly_name:Contig93"
223964. .230362
/note="assembly_name:Contig92"
230383. .236251
/note="assembly_name:Contig91"

Qy	2085	gaaagaagtgggaagagggtacctgcagaagtagtaagaagtgggaagagggtaccctgcagaagta	2144
Db	177896	CGAGGAGGAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	177837
Qy	2145	gaaagaagtgggaagagggtaccagaagaagttagaagaaggttaccctgcagaagtagaagaagtg	2204
Db	177836	GGACGAGGAGCAGCAGGAACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	177777
Qy	2205	gaagaggtaccagaagaagtgggaagaggtaccagaagaagtgggaagaggtaccagaagaag	2264
Db	177776	GGAGGAGCAGGAGCAGCAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGCAGGAGCA	177717
Qy	2265	gtggaagagggtaccagaagaagtgggaagaagtgggaagaagttagaagaagtagaagttacca	2324
Db	177716	GCAGCAGGAGGAGCAGGAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCA	177657
Qy	2325	gcggtagtagaagttagaaggtaccacgcggttagaagaagaaggtgccagaagaagtagaag	2384
Db	177656	GCAGGAGCAGGAGCAGCAGCAGGAGGAGCAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCACTGGGGCA	177559
Qy	2385	gaagaacaagaaggaagaaccactagtagaagaagaatata	2426
Db	177596	GGACAGGAGCAGAGAAGGGCGGGGCCAGGAGTGGGATGGGA	177555

RESULT 15
AC018634/c

LOCUS AC018634 174241 bp DNA linear PRI 28-JUL-2000

DEFINITION Human Chromosome 7 clone RP11-243E12, complete sequence.

ACCESSION AC018634

VERSION AC018634.3 GI:9558607

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174241)
Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and Olson,M.V.

AUTHORS Large-scale Mapping and Sequencing of Human Chromosome 7 Unpublished

TITLE 2 (bases 1 to 174241)

JOURNAL Bubbs,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.

AUTHORS Direct Submission

TITLE Submitted (15-DEC-1999) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

JOURNAL Washington, Box 352145

REFERENCE 3 (bases 1 to 174241)

AUTHORS Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and Olson,M.V.

TITLE Direct Submission

JOURNAL Submitted (28-JUL-2000) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

COMMENT On Jul 28, 2000 this sequence version replaced gi:8050915.

----- Genome Center

Center: University of Washington Genome Center

Center Code: UWGC

Web site: <http://www.genome.washington.edu/UWGC/>

Contact: uwgchtgsu.washington.edu

----- Project Information

Center project name: HsaChr7

Center clone name: RP11-243E12 (djsl56)

----- Summary Statistics

Sequencing vector: M13; X02513; 100% of reads

Chemistry: Dye-primer-amersham; 66% of reads

Chemistry: Dye-terminator ET-amersham; 34% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 173493 bases at least Q40

Consensus quality: 174139 bases at least Q30

Consensus quality: 174231 bases at least Q20

Insert size: 179271; 8.6% error; agarose-fp

Insert size: 174241; sum-of-contigs

Quality coverage: 6.32x in Q20 bases; agarose-fp

Quality coverage: 6.50x in Q20 bases; sum-of-contigs

 Overlapping Sequences:
 5': RP11-605P22 (UMGC:djs708)
 3': Mapping in progress

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.
 Double stranded (DS) coverage: 70.2%
 DS or two chemistry coverage: 29%
 Single stranded regions: 0.787% in 7 gaps

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below.
 The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BgIII				EcoRI				NsiI			
SeqDerMap	FngrPrint	SeqDerMap	FngrPrint	SeqDerMap	FngrPrint	SeqDerMap	FngrPrint	SeqDerMap	FngrPrint	SeqDerMap	FngrPrint
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
2554	2514	8696	8906	12972	12987						
-----	-----	-----	-----	-----	-----						
2067	2002	210	<800	1397	1379						
-----	-----	-----	-----	-----	-----						
4648	4783	605	<800	4332	4323						
-----	-----	-----	-----	-----	-----						
3573	3614	545	<800	6863	6887						
-----	-----	-----	-----	-----	-----						
8078	8030	4828	4820	3199	3170						
-----	-----	-----	-----	-----	-----						
4368	4326	774	782	1887	1838						
-----	-----	-----	-----	-----	-----						
16815	16889	11779	11746	1024	1024						
-----	-----	-----	-----	-----	-----						
2290	2300	2846	2851	4955	4937						
-----	-----	-----	-----	-----	-----						
4064	4019	7395	7450	2501	2457						
-----	-----	-----	-----	-----	-----						
12849	12884	15587	15482	167	<800						
-----	-----	-----	-----	-----	-----						
3693	3850	4041	4019	4213	4186						
-----	-----	-----	-----	-----	-----						
2740	2714	3827	3801	7678	7631						
-----	-----	-----	-----	-----	-----						
3872	4019	3925	3916	3405	3413						
-----	-----	-----	-----	-----	-----						
491	<800	2972	2950	2335	2385						
-----	-----	-----	-----	-----	-----						
1890	1885	1716	1728	1344	1327						
-----	-----	-----	-----	-----	-----						

6066	5966	18105	18158	1076	1043
-----	-----	-----	-----	-----	-----
867	823	4047	4019	467	<800
-----	-----	-----	-----	-----	-----
7550	7595	479	<800	938	991
-----	-----	-----	-----	-----	-----
1542	1531	7518	7450	3184	3170
-----	-----	-----	-----	-----	-----
5259	5070	115	<800	2431	2385
-----	-----	-----	-----	-----	-----
2061	2053	10695	10628	130	<800
-----	-----	-----	-----	-----	-----
3357	3364	1142	1121	2209	2189
-----	-----	-----	-----	-----	-----
7394	7324	4194	4160	2227	2231
-----	-----	-----	-----	-----	-----
813	823	3104	3123	1857	1838
-----	-----	-----	-----	-----	-----
5190	5070	1019	1012	8477	8468
-----	-----	-----	-----	-----	-----
308	<800	9558	9461	393	<800
-----	-----	-----	-----	-----	-----
4317	4326	976	973	1041	1043
-----	-----	-----	-----	-----	-----
2748	2714	1129	1121	139	<800
-----	-----	-----	-----	-----	-----
6726	6665	10185	10126	821	816
-----	-----	-----	-----	-----	-----
6831	6877	1487	1486	3610	3669
-----	-----	-----	-----	-----	-----
2355	2393	433	<800	618	<800
-----	-----	-----	-----	-----	-----
2172	2140	3371	3385	4153	4186
-----	-----	-----	-----	-----	-----
1916	2002	10484	10628	3167	3170
-----	-----	-----	-----	-----	-----
2396	2514	10143	10126	630	<800
-----	-----	-----	-----	-----	-----
2512	2514	8539	8577	2314	2231
-----	-----	-----	-----	-----	-----
4092	4326	663	<800	1846	1800
-----	-----	-----	-----	-----	-----
900	920	122	<800	3624	3669
-----	-----	-----	-----	-----	-----
1618	1614	174	<800	1026	1024
-----	-----	-----	-----	-----	-----
6696	6534	4551	4590	807	816
-----	-----	-----	-----	-----	-----
4460	4326	952	973	190	<800
-----	-----	-----	-----	-----	-----
5469	5309	7	<800	1808	1800
-----	-----	-----	-----	-----	-----
3387	3614			987	991
-----	-----	-----	-----	-----	-----
2195	2300			1806	1800
-----	-----	-----	-----	-----	-----
2555	2514			513	<800
-----	-----	-----	-----	-----	-----
1416	1402			4060	4186
-----	-----	-----	-----	-----	-----
610	<800			398	<800
-----	-----	-----	-----	-----	-----
225	<800			274	<800
-----	-----	-----	-----	-----	-----
932	920			983	991
-----	-----	-----	-----	-----	-----
1982	1885			722	726
-----	-----	-----	-----	-----	-----
29	<800			1638	1760
-----	-----	-----	-----	-----	-----
74	<800				
-----	-----	-----	-----	-----	-----
853	844				

Db 1870 AGTTAAAAAAGTGTAAAAAAGTAAAAAAGGAAATTAAGCAAAATGATCTGAAGG 1929
Qy 548 tggcgatatgaagcagggaagaagctggtgatttggagcaggggaagaactgcgcga 607
Db 1930 TAACGATAAAGTGAAGAGCACGAGAAATAATTAATGAAGAGTAAAGAGCAAAATTTAAAA 1989
Qy 608 ttggagcagggggaagaactgcgcgatttggagcaggggaagaagctggtgatttggga 667
Db 1990 ACAAGTAGAAGATGGAATTAAGAAAAATGATCTGAAGCTTAACGATAAAGTGAAGGGCC 2049
Qy 668 agcagggggaagaactgcgcgatttggagcaggggaagaactgcgcgaaactga 727
Db 2050 AGAAATTAATCTCGAAGAGTAAAGAGAAATTTAAAAACATCTGAAGAGGTATTA 2109
Qy 728 agaaggcgcaactgagatgcgaaactgaaatgg-----agcaactgtgataga 781
Db 2110 AGAAATTAATCTCAAGGTAAAGTAAAGTCAAAAGGCCAGAAATAATACTGAAGAGT 2169
Qy 782 cacagaagatagttcagctgatggcagaaaaagttacatgttcttgcctcgaagaatgt 841
Db 2170 AAAAGAGAAATTAAGAAACATGTAGAAGAGGAATTAAGAAATGATCTGAAAGTAA 2229
Qy 842 acaactgcgcgatagtaatatgccctcttggagttatttggataaaatataat 901
Db 2230 AGATAAAGTGTAGGCAAGAAATAATACTGAAGAGTAAAGAGGAATTTAAAGAAAA 2289
Qy 902 tgatcatatlaagatttcgagccactattcgaacaaattgtgcggttactgctaaaca 961
Db 2290 TGATCTGAAAAATAAGATAAAGTGTAGGACCAAGTAATTAACGAAAGTAAAAA 2349
Qy 962 tgttacgggacaagaattgccaatgaacctgtaccattaccgtggcagaagccgc 1021
Db 2350 AGAAATTTGAAAAACAGAGAAAGAAATAAAGAAATATTCTTGAAATTTAAAGATAT 2409
Qy 1022 gcaagtaccgcggaagaattagatgccactccagagatgacttcgattagattac 1081
Db 2410 AGTAATTTGGACAGAGTAAATATAGAAAGTAAATAAAGTAAATAAATAAAGTAGA 2469
Qy 1082 agaattcccgaggaagtagaattagtagtagagagcgcaactgaagaactcaac 1141
Db 2470 AAAAGGAATTAAGAAAAATCATCTGAAAGTAAAGATAAAGTAAAGTAAAGATAAT 2529
Qy 1142 ggaagtgggaccacggaagaagaccacccgaagaattagatgccactccagagatgg 1201
Db 2530 AGTTGAAGAGTAAAGAGAAATTTGAAACCAAGTAGAAGAGGAATTTAAAGAAATGA 2589
Qy 1202 atttcgattagacgaactgcgaagagagaacagagaacgtagagggagagaac 1261
Db 2590 TACTGAAAGTAAAGATAAAGTGTAGGACAAAGTATTAAGGAGATGTTAATGAAGA 2649
Qy 1262 agaagaagctgcgaaggagaagtagatcagaagaactccagaaggaggaagagtaga 1321
Db 2650 AGGTCCGAAAAACAAAGATAAAGTACAAAAACAGGAAAAAGTAAAGAACTTAAAAAAGA 2709
Qy 1322 ggaactccagaggaatgatttcgattagatggaactcaattagaagaaccgaagaac 1381
Db 2710 AGTAAAAAAGAGTTAAAAAAGAGTAAA-----AAAAAGAAATTAATGAAGTGAAGAAA 2766
Qy 1382 tgcagaaggagaagaaccctagaggagagaaccctagaggaggaagaacccttaga 1441
Db 2767 AGATAATGTTATAGGAAAGAAATAATGAAGAGAGTGTAAATGAAGAAATACCCGAAA 2826
Qy 1442 ggaagaagaagctgcagaaggaagaagatttagaggccaactccagagatgacttcca 1501
Db 2827 CAAGATAAAGAGATAGAACAAAAAGAAAAAGAAAGAGTAAAGAAAAAGAAAGTAAAGA 2886
Qy 1502 attagaagaaccatcagagaaggaaggggaaggaaggaaggaaggaaggaaggaag 1561
Db 2887 TAAAGAAAAAGAGAGTAAAGAAAAAGAAAGAAAGTAAAGAAAAAGAAAGTAAAGA 2946
Qy 1562 agaagcgttagtagcagtagtggccgaaccggttagaagtagtactctctgctca 1621
Db 2947 AAAAGAGAGTAAAGAAAAAGAAAGTAAAGAAAAAGAAAGTAAAGAAAAAGAA 3006

Qy 1622 gcctgtcaaaccaatgctcgctcccaacggcagatgaacttttattgattgatatcttaga 1681
Db 3007 AGAAGTAAAAAAGAAAAAGATACCGAAAGCAAAAGATAAAGAGATAGAACAAACAAAAAGAAA 3066
Qy 1682 taacgatttaacgtatgcagacattacatcttggcgccattttaaacaatctctcaa 1741
Db 3067 AGAAGAGTAAAAAAGAGTTAAAGAAAAAGATACCGAAAAAAGTAAAGTATAGGACA 3126
Qy 1742 ggaactgtatgcagagagcgtgtaacagtaccatcaaaaggaagcaccctgtacaagtacc 1801
Db 3127 AGAAATTAATTAAGAGAAATAAATAAAGAGTAAATAAAGAGTAAAAAAGAAATAA 3186
Qy 1802 agtggcagtagggccgcgcgaagtgcccaacggaagaattgtgcaactcccaaggga 1861
Db 3187 TAAAAATGAACCAAAAGATAATGTGTAGTACAGAAATAATGAACGAGATGTTACGAAA 3246
Qy 1862 cgatttcgaattagaaggaaactgcagaagctccagagggaagggaatttagtattagaagg 1921
Db 3247 AGATACCGCAAAACAAAGATTAAGGTGATAGAACAAAGAAAAAGAAAAAGAAAGTAAAAA 3306
Qy 1922 agaagagaacccaacggaagaagagccaagagaagagagccaacagagaaggaagtgc 1981
Db 3307 AAAAGAGAGTAAAGAAAAAGAAAGTAAAGAAAAAGAAAGTAAAGAAAAAGAAATAA 3366
Qy 1982 agaagaagaattagaggaactccagaggcagatttcgaattagaagaacccaacagaga 2041
Db 3367 AGAAGTAAAAAAGAAAAAGATACCGAAAGCAAAAGATAAAGAGATAGAACAAAGAAAAA 3426
Qy 2042 agaagttagaagaacccgttagaggcggaagaactgcagaagggaagaactggaagaggt 2101
Db 3427 AGAAGAGTAAAAAGAGTTAAAGAAAAAGATACCGAAAAAAGTAAAGTATAGGACA 3486
Qy 2102 acctgcagaagtagaagaagtggaagaggtcacctgcagaagtagaagaagtggaagaggt 2161
Db 3487 AGAAATTAATTAAGAGAAATAAATAAAGAGTAAATAAAGAGTAAAAAAGAAATAA 3546
Qy 2162 accagaagaagtagaagaggtaccgcagagtagaagaagtggaaagaggtaccagaaga 2221
Db 3547 TAAAAATGAACCAAAAGATAATGTGTAGTACAGAAATAATGAACGAAAGTGTATACGA 3606
Qy 2222 agtggaaaggttaccagaagaagtgaaagaggtaccagaagaagtggaagaggtaccaga 2281
Db 3607 AAAAGTATACCGCAAAACAAAGATTAAGGTGATAGAACAAAGAAAAAGAAAGAAAGTAAA 3666
Qy 2282 agaagtggagaagtggaagaagtagaagaagtagaggttaccgcggttagtagaagtaga 2341
Db 3667 AGAAAAAGAGAGTAAACAAAAAGAAAGTAAAGAAAAAGAAAGAAAGTAAAAAGAAA 3726
Qy 2342 agtaccagcggtagttagaagaaggtgcccagaagaagtagagaagaagaagaag----- 2394
Db 3727 AGATACCGAAAGCAAAAGTAAAGAGATAGAACAAAGAAAAAGAAAAAGAAAGTAAAAAGA 3786
Qy 2395 -----aagaggaagaaccagtagagaagaagtagtattacaattagtagtaaccatcgga 2449
Db 3787 AGTTAAAGAAAAAGATACCGAAAAACAAAGATAAAGTGTAGGACAAAGAAATAATATAGA 3846
Qy 2450 agaagatatcaattagacaacccaagaagagcgaattaggctctggaattttatctat 2509
Db 3847 AGAAATAAAAAAGAGTTAAAAAAGAGTAAAAAAGAAATAATAAAAATGAACAA 3906
Qy 2510 catcgacatgcactaccaagacgttccaaaggaa---tttatggagaagaagaagaac 2566
Db 3907 AGATAATGTGTATAGTACAGAAATAATGAACGAAAGTGTAAACGAAAAAGATACCCGAAA 3966
Qy 2567 tgcagtgattcattgaaaccagaagatttgcagaagaagattccacaattctacagaagt 2626
Db 3967 CAAGATAAGGTGTAGAACAAAGAAAAAGAAAGAGAGAGTAAAGAAAGTAAAGAAAA 4026
Qy 2627 gctcacattcatccaagggcctagaadgcgactgggaacgattagaagtagagcttaataa 2686
Db 4027 AGATACCGTAAACAAAGATAAAGTGTAGGACAAGAGTAAATAAGAAAGTAAAAAGA 4086


```
* 51200 53719: contig of 2520 bp in length
* 53720 53819: gap of 100 bp
* 53820 56012: contig of 2193 bp in length
* 56013 56112: gap of 100 bp
* 56113 57226: contig of 1114 bp in length
* 57227 57326: gap of 100 bp
* 57327 77403: contig of 20077 bp in length
* 77404 77503: gap of 100 bp
* 77504 79994: contig of 2491 bp in length
* 79995 80094: gap of 100 bp
* 80095 82923: contig of 2829 bp in length
* 82924 83023: gap of 100 bp
* 83024 86207: contig of 3184 bp in length
* 86208 86307: gap of 100 bp
* 86308 90457: contig of 4150 bp in length
* 90458 90557: gap of 100 bp
* 90558 96908: contig of 6351 bp in length
* 96909 97008: gap of 100 bp
* 97009 102861: contig of 5853 bp in length
* 102862 102961: gap of 100 bp
* 102962 107881: contig of 4920 bp in length
* 107882 107981: gap of 100 bp
* 107982 113543: contig of 5562 bp in length
* 113544 113643: gap of 100 bp
* 113644 118163: contig of 4520 bp in length
* 118164 118263: gap of 100 bp
* 118264 135271: contig of 17008 bp in length
* 135272 135371: gap of 100 bp
* 135372 151730: contig of 16359 bp in length
* 151731 151830: gap of 100 bp
* 151831 171146: contig of 19316 bp in length
* 171147 171246: gap of 100 bp
* 171247 174019: contig of 2773 bp in length.
```

FEATURES

source

```
1. .174019
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="15"
  /map="15"
  /clone="RP11-228G18"
  /clone_lib="RPC1-11 Human Male BAC"
```

misc_feature

```
1. .13487
  clone_end:SP6
  vector_side:left"
  /note="assembly_fragment"
```

misc_feature

```
13588..14626
  /note="assembly_fragment"
```

misc_feature

```
14727..15793
  /note="assembly_fragment"
```

misc_feature

```
15894..17057
  /note="assembly_fragment"
```

misc_feature

```
17158..18291
  /note="assembly_fragment"
```

misc_feature

```
18392..20046
  /note="assembly_fragment"
```

misc_feature

```
20147..21386
  /note="assembly_fragment"
```

misc_feature

```
21487..22760
  /note="assembly_fragment"
```

misc_feature

```
22861..23953
  /note="assembly_fragment"
```

misc_feature

```
24054..25081
  /note="assembly_fragment"
```

misc_feature

```
25182..26357
  /note="assembly_fragment"
```

misc_feature

```
26458..27574
  /note="assembly_fragment"
```

misc_feature

```
27675..28756
  /note="assembly_fragment"
```

misc_feature

```
28857..30554
  /note="assembly_fragment"
```

misc_feature

```
30655..32124
  /note="assembly_fragment"
```

```
misc_feature 32225..33634
              /note="assembly_fragment"
misc_feature 33735..35520
              /note="assembly_fragment"
```

```
Query Match 7.7%; Score 255.8; DB 2; Length 174019;
Best Local Similarity 48.6%; Pred. No. 7.6e-28;
Matches 918; Conservative 0; Mismatches 942; Indels 27; Gaps 7;
```

```
QY 531 gaaggaagaagaactgctgcataatgaagcaggaggaagaagctggtggaagca 590
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66198 GGAGGAGATGTGGAGCAGGAGGAGGAGCTACGGGACGAGAAAGAAATACGGAAGCA 66257
QY 591 ggggaagaactggcgatttgggaagcagggaagaactggcgatttgggaagcaggggaa 650
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66258 GGAGGAGAA-----GATGTGGAGACAGGAGGAGAGACTCGG---GAGCAGGAGGGGAA 66308
QY 651 gaagctgggtgatttgaagcagggggaagaactggcgatttgaagcagggggaagaact 710
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66309 GATGCGGGAGCAGGAGGAGGAGAAAGATGCGGAGACAGAGAGAGAGCTCGGAGAGGAGAA 66368
QY 711 ggagatgcggaactgaagaagagcaactggagatgcggaactggaactggaactgagcaact 770
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66369 GGAGCTCGGGGACGAGGAGGAGGAGCTGCGGGAGCAGAAAGCTCGGGGAGCAGGAGGA 66428
QY 771 gtgtatgtagacacagaagatagttcagctgtatggagcagaaaaagtacatgttctcgt 830
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66429 GCAGATGCAGGAGCAGGAGGAGGAGAAAGATGTGGGAGCAGGAGGAGAAAGATGCGGAGCAAGCA 66488
QY 831 caagaataatgacaactgccgatagtaatgatgcctcttggaaagtatttggataaa 890
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66489 AGAGAGATGTGGAGACAGGAGGAGAGGCTGTGGGAGCAGGAGAAAGCAGATGCGGGAGCA 66548
QY 891 gatataattttgatcatattaaagatttcgagcactattcgaaacaaattgtggcgggt 950
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66549 GGAGCAGAGAGATCGGGACCAACAGAGAGAGGATGTGGGAGCAGGACAGAGAGGTACGGGA 66608
QY 951 actgctaaacattgttacgggacaagaattgccaatgaaacc---tgtaccattaccagt 1007
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66609 GAAGGAGGAGAGGATCGGGAGCAGGAGGAGAAAGATGTGGGAGCAGGTTGGAGAAAGATGCGGGGA 66668
QY 1008 gcagaagagccgcgcgaagtaccagcggaagaatttagatccactccagaggatgacttc 1067
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66669 GGAGAGAGAGATGCAGGAGCAGGAGAGAGAGACGCCGGACAGGAGGAGAGATGCAAGA 66728
QY 1068 gcattagattttacagaatctcccgaggaa---gtagaattagttattagatgaagggca 1124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66729 GGAGGAGAGGATACGGGAGCGGGAGAGAGAGATCGGGGAAAGAGAGGAGAGACGATGCGGGA 66788
QY 1125 actgaagaagaatcaacggaagtgggaccacaaggaagaagaccacccgaagaattagat 1184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66789 GCAGGAGGAGAGATGCAGAAAGCAGGAGGAGAAATATGTGGGAGCAGGAGAAAGAGAGTGTG 66848
QY 1185 gccactccagagagatttcgatttcattagacgaactgcagaaggaagaacacagaagaac 1244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66849 GCAGCAGCAGAGGCTCCCGGAACAGAGAGGAGAGAGCTGTGGGAAACAGGAGAGATGCAAGA 66908
QY 1245 gttagaggagaagaacacagaagaagctgcagaagggagaagatatcagaagaactcccgaa 1304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66909 GCAGGAGGAGAGATATGTGGGAGCAGGAGGAGAGAGATACGCGGACCGAGGAGAGATGTGGGG 66968
QY 1305 ggagaagaagtttagagcaactccagaggtatttcgca-ttagatggaactacatt 1363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66969 GCAGGAGAGAGAGATGTGGCGGCGGAGGAGAGATGCGGGAGCAGGAGAAAGATGTGGAGACAGG 67028
QY 1364 agaagaacccaagaactgcagaaggaagaacacgtagagggaagaagaacacgtaga 1423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67029 AGGAGAGGCTGCAGGAGCAGGAGAGAGAGATGTGGGAGCAGGAGGAGAGAGATGCGGGATC 67088
QY 1424 gggagaagaacacgct---agagggaagaagctgcagaagggagaagaagattagagggc 1480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67089 AGGAGCAGAAAGATGTGGGACCGAGGAGAGGATGTGGGAGCAGGAGGAGGAGGATGCGGGG 67148
```

QY	1481	aactccagaagtagtacttccaattagaagaaccatcaggagaagggaggaagggggaagaga	1540
Db	67149	AGAAAGGAGGAGAGAATGCGGGAGCAGGAAGAAGATGTGGCAGCAGAGTGCAGAAATGCGGG	67208
QY	1541	aggagaaggggaaggaaggaacgcttagtcagcagtcgcaagtagtgccgaacccggt	1600
Db	67209	AGAGAAGAAGACGCGAGGACGAGGAAGAAGACATGGGACCAGGAGGAGAAGATGCGGAG	67268
QY	1601	agaagtagtgactcctctgcagcctgtccaaaccaatggctcgctcccaacgagcagatgaac	1660
Db	67269	AGGAGGAGGAGATGCGGGAGCGGAGGAAGAAGATGCGGGAGGAGGAGGAGATGCGGG	67328
QY	1661	tttatctgtgatactcttagataacagatttaacgtagtcagacaattacatccctttgagcc	1720
Db	67329	AGCAGGAGGAGAAGATGCGAGGACGAGGAAGAAGATGCGGGAGCAGGAGGAGAAGATGT	67388
QY	1721	attatttaacaaatcctcaagatcctgatcagagagagcgtgtacacgtaccatcaaa	1780
Db	67389	GGAGGAGGAAGAAGATGCGAGGACGAGGAGGAGAAGATGTGGGACGAGGAGGAAGA	67448
QY	1781	ggaagcacctgtacaagtccagtggcagtagggcccgccgaagaagtgccaaacggaaga	1840
Db	67449	TGTGGGAGCAGGAAGAAGATCTGGGAGCAGCAGAGCGCTGCCGGAACAGAGAGCAGC	67508
QY	1841	attgatgcaactccaagaggacg-----atttcgaattagaagaactgcagaagct	1892
Db	67509	TGTGGGAACACGAGAAGATCGAGGACGAGGAGATCAGGAGCAGAAATGTGGGACCAGAA	67568
QY	1893	ccagagaaggaagaattagattagaaggaaggaacccaacggaagaagaccaga	1952
Db	67569	GGAGAGGATGTGGGACGAGCAGCAGAGCGTCCGGGAGAGGAGGAGGAATGCGGGAGCA	67628
QY	1953	gaaggagagccaagaagaagtgccagagaagaataagaggcaactccagaggac	2012
Db	67629	GAAGATGTGCACAGGTGGAGGAAGATGCGGGAGAGAGAAACACCGCAGGAGGAAGAT	67688
QY	2013	gatttcgaattagaagaaccaacagagagaagaagtagaagaacccgtagagggcgaaaga	2072
Db	67689	GCAGGAGCAGGAAGAAGATGCGAGGACGAGGAGGAGATGTGGGAGCAGGAGGAAGATG	67748
QY	2073	actgcagaaggaagaagtggaagaggtaccagaagaagtagaagaaggtaccgcgagaa	2132
Db	67749	GATGCGAGGACGAGGAGGAAGATGTGGGAGCAGGAGGAGGAAGATGCGGGAGCAGGAGGA	67808
QY	2133	cctgcagaagtagaagaagtgggaagaggtaccagaagaagtagaagaaggtaccgcgagaa	2192
Db	67809	GAAGATGCGGAGCAGGAGGAGAAGATGCGAGGCCAGGAGGAGCAAGATGCGGGAGCAGGA	67868
QY	2193	gtagaagaagtggagaaggtaccagaagaagtggaagaggtaccagaagaagtggaagag	2252
Db	67869	GGAGAAGATGCGGGCCATGAGGAGAAGATGCGGGAGCAGGAGGAGAAGATGCGGGGCCA	67928
QY	2253	gtaccagaagaagtgaagaaggtaccagaagaagtggaagaagtggaagaagtagaaga	2312
Db	67929	GGAGGAGAAGATGTGGGCCAGGAGGAGAAGATGTGGGCCAGGAGGAGAAGATGTGGGG	67988
QY	2313	gtagaggtaccagcggttagtagaagtagaagtagtaccagcggttagtagaagaaggtgcc	2372
Db	67989	GCAGGAGGAGATCGAGAGAAGGAGGAGGAGGATACGAGATCAACAAGAGAAGATGCGAGGA	68048
QY	2373	gaagaagtagaagaagaagaagaagag	2399
Db	68049	GAGGCTGCCAGACGAGGAGCGGTG	68075
RESULT	22		
AC106251			
LOCUS	AC106251	170276 bp	DNA
DEFINITION	Rattus norvegicus clone CH230-201A10, *** SEQUENCING IN PROGRESS		
	***, 62 unordered pieces.		
ACCESSION	AC106251		
VERSION	AC106251.1	GI:18138772	
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
REFERENCE	1 (bases 1 to 170276)		
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Allbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowles,S., Brivia,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Checko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dunn-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsif.F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,M., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshkari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 170276)		
AUTHORS	Worley,K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
COMMENT	----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GKHO Center clone name: CH230-201A10 ----- Summary Statistics Assembly program: Phrap; version 0.990329First call to findPhrapList Consensus quality: 139818 bases at least Q40 Consensus quality: 147078 bases at least Q30 Consensus quality: 153529 bases at least Q20 Estimated insert size: 145171; sum-of-contigs estimation Quality coverage: 0x in Q20 bases; agarose-fp estimation Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).		

1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

EcoRI				HindIII				BglII			
SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8696	8850	399	<800	2304	2280						
-----	-----	-----	-----	-----	-----						
6	<800	6382	6347	2067	2069						
-----	-----	-----	-----	-----	-----						
1907	1895	512	<800	5518	5637						
-----	-----	-----	-----	-----	-----						
1854	1895	449	<800	6918	7002						
-----	-----	-----	-----	-----	-----						
2849	2922	2322	2276	2029	2069						
-----	-----	-----	-----	-----	-----						
3722	3806	5951	5995	2378	2425						
-----	-----	-----	-----	-----	-----						
658	<800	7224	7291	2089	2069						
-----	-----	-----	-----	-----	-----						
26398	26593	1005	1033	550	<800						
-----	-----	-----	-----	-----	-----						
537	<800	1020	1033	4191	4187						
-----	-----	-----	-----	-----	-----						
1542	1483	3445	3448	3839	3773						
-----	-----	-----	-----	-----	-----						
1103	1096	2131	2126	1647	1646						
-----	-----	-----	-----	-----	-----						
1928	2045	2237	2276	629	<800						
-----	-----	-----	-----	-----	-----						
14029	13827	6625	6652	1932	1918						
-----	-----	-----	-----	-----	-----						
6707	6816	6349	6347	18266	18041						
-----	-----	-----	-----	-----	-----						
1884	1895	6669	6652	3099	3165						
-----	-----	-----	-----	-----	-----						
6048	6029	1943	1916	2900	2928						
-----	-----	-----	-----	-----	-----						
2150	2209	6376	6347	111	<800						
-----	-----	-----	-----	-----	-----						
13944	13827	5398	5641	2425	2425						
-----	-----	-----	-----	-----	-----						
4275	4287	599	<800	3767	3773						
-----	-----	-----	-----	-----	-----						
2057	2209	4055	4127	1506	1485						
-----	-----	-----	-----	-----	-----						
2293	2209	299	<800	3639	3773						
-----	-----	-----	-----	-----	-----						
6448	6393	20065	20079	5635	5637						

FEATURES

source
Location/Qualifiers
1. .179149
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="Rp11-198E23"
/clone_lib="RPC1 human BAC library 11"
complement(3542..3629)
/standard_name="SWSS3607"
complement(87234..87324)
/standard_name="SWSS3670"
115832..116064
/note="Single subclone region"
130498..130537
/note="Single subclone region"

STS

STS

misc_feature

misc_feature


```

AC098094/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-11K6, *** SEQUENCING IN PROGRESS ***
68 unordered pieces.
AC098094.4 GI:17968682
HTG: HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 157158)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Slisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 157158)
Worley,K.C.
Direct Submission
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062340.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGC1
Center clone name: CH230-11K6
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 129931 bases at least Q40
Consensus quality: 136277 bases at least Q30
Consensus quality: 140929 bases at least Q20
Estimated insert size: 122547; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5265: contig of 5265 bp in length
* 5266: gap of unknown length
* 11219: contig of 5854 bp in length
* 11220: gap of unknown length
* 11320: contig of 3052 bp in length
* 14371: gap of unknown length
* 14372: gap of unknown length
* 14472: contig of 4616 bp in length
* 19087: gap of unknown length
* 19088: contig of 4269 bp in length
* 19188: gap of unknown length
* 23457: contig of 3975 bp in length
* 27531: gap of unknown length
* 27532: contig of 4150 bp in length
* 31781: gap of unknown length
* 31782: gap of unknown length
* 36040: contig of 4159 bp in length
* 36041: gap of unknown length
* 39064: contig of 2924 bp in length
* 39164: gap of unknown length
* 39165: contig of 3792 bp in length
* 42956: gap of unknown length
* 43057: gap of unknown length
* 46114: contig of 3558 bp in length
* 46114: gap of unknown length
* 46715: contig of 2890 bp in length
* 49604: gap of unknown length
* 49605: contig of 3132 bp in length
* 52836: gap of unknown length
* 52937: contig of 3222 bp in length
* 56158: gap of unknown length
* 56258: gap of unknown length
* 58770: contig of 2612 bp in length
* 58870: gap of unknown length
* 61840: contig of 2870 bp in length
* 61841: gap of unknown length
* 65240: contig of 3300 bp in length
* 65241: gap of unknown length
* 67380: contig of 2040 bp in length
* 67480: gap of unknown length
* 70683: contig of 3203 bp in length
* 70684: gap of unknown length
* 73008: contig of 2225 bp in length
* 73009: gap of unknown length
* 75331: contig of 2223 bp in length
* 75332: gap of unknown length
* 78297: contig of 2865 bp in length
* 78297: gap of unknown length
* 80629: contig of 2233 bp in length
* 80630: gap of unknown length
* 83631: contig of 2902 bp in length
* 83731: gap of unknown length
* 86848: contig of 3117 bp in length
* 86849: gap of unknown length
* 88358: contig of 1410 bp in length
* 88359: gap of unknown length
* 90496: contig of 2037 bp in length
* 90496: gap of unknown length
* 93201: contig of 2606 bp in length
* 93202: gap of unknown length
* 95328: contig of 2027 bp in length

```


ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP11-328C17 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-328C17 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-328C17 is at 132449 in this sequence. The true left end of clone RP11-233K4 is at 64893 in this sequence. The true right end of clone RP3-416J7 is at 100 in this sequence. The true right end of clone RPI-125A24 is at 61243 in this sequence.

FEATURES

Location/Qualifiers

Source

1..132449
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-328C17"
/clone_lib="RPCI-11.2"

repeat_region
1..1562
/note="LIM4 repeat: matches 4133. .5719 of consensus"
repeat_region
1563..1641
/note="LIM1A2 repeat: matches 1. .80 of consensus"
repeat_region
1669..2199
/note="LIM4 repeat: matches 3396. .3937 of consensus"
repeat_region
2200..2474
/note="LIM4 repeat: matches 38. .312 of consensus"
repeat_region
2475..4279
/note="LIM4 repeat: matches 1686. .3396 of consensus"
repeat_region
4306..4589
/note="LIM4 repeat: matches 1254. .1544 of consensus"
misc_feature
complement(4704..5324)
/note="match: GSS: Em:AQ628183"
repeat_region
4799..4938
/note="LIM4c repeat: matches 785. .924 of consensus"
misc_feature
complement(4809..5326)
/note="match: GSS: Em:AQ774852"
misc_feature
complement(4860..5272)
/note="match: GSS: Em:AQ787689"
repeat_region
4943..5094
/note="LIM4c repeat: matches 272. .420 of consensus"
misc_feature
complement(5020..5329)
/note="match: GSS: Em:AQ494601"
repeat_region
5142..5435
/note="LIM1D repeat: matches 170. .428 of consensus"
repeat_region
5436..5744
/note="AluX repeat: matches 1. .310 of consensus"
repeat_region
5745..5869
/note="LIM1D repeat: matches 428. .568 of consensus"
repeat_region
6450..6686
/note="MIR repeat: matches 7. .261 of consensus"
repeat_region
6916..7994
/note="LIM4 repeat: matches 3582. .4688 of consensus"
repeat_region
7984..8050
/note="LIMB8 repeat: matches 5808. .5884 of consensus"
repeat_region
8064..8169
/note="LIM4 repeat: matches 3479. .3593 of consensus"
repeat_region
8178..8224
/note="LIMB4 repeat: matches 6124. .6169 of consensus"
repeat_region
8225..8525
/note="AluX repeat: matches 1. .301 of consensus"
repeat_region
8526..8915
/note="LIMB4 repeat: matches 5725. .6124 of consensus"

misc_feature
complement(8862..9353)
/note="match: GSS: Em:AQ355497"
repeat_region
9029..9150
/note="2 copies 61 mer 95% conserved"
misc_feature
complement(9052..9484)
/note="match: GSS: Em:AQ175845"
misc_feature
9388..9675
/note="match: GSS: Em:AQ135206"
misc_feature
9487..9840
/note="match: GSS: Em:AQ171684"
misc_feature
9491..9932
/note="match: GSS: Em:AQ147759"
repeat_region
9524..9697
/note="MER53 repeat: matches 1. .189 of consensus"
repeat_region
10046..10143
/note="2 copies 49 mer 91% conserved"
repeat_region
11168..11361
/note="MT1B repeat: matches 101. .309 of consensus"
repeat_region
11448..11736
/note="AluY repeat: matches 1. .309 of consensus"
repeat_region
12109..12515
/note="L2 repeat: matches 1747. .2155 of consensus"
repeat_region
12798..12947
/note="MIR repeat: matches 29. .206 of consensus"
repeat_region
12975..13026
/note="MIR repeat: matches 186. .237 of consensus"
repeat_region
13005..13103
/note="L2 repeat: matches 2417. .2516 of consensus"
repeat_region
13331..13664
/note="MT1B repeat: matches 4. .361 of consensus"
repeat_region
13704..13907
/note="LIR33 repeat: matches 297. .511 of consensus"
repeat_region
15450..15530
/note="L2 repeat: matches 2668. .2749 of consensus"
repeat_region
15876..15935
/note="2 copies 30 mer 100% conserved"
repeat_region
16202..16234
/note="MER81 repeat: matches 2. .94 of consensus"
misc_feature
17335..17668
/note="match: GSS: Em:AQ104665"
repeat_region
17419..17715
/note="AluX repeat: matches 1. .296 of consensus"
misc_feature
17995..18362
/note="match: GSS: Em:AQ151578"
repeat_region
19374..19496
/note="L2 repeat: matches 2579. .2709 of consensus"
repeat_region
20271..20430
/note="MER20 repeat: matches 55. .218 of consensus"
repeat_region
21095..21389
/note="AluX repeat: matches 1. .295 of consensus"
repeat_region
21830..21946
/note="MIR repeat: matches 42. .152 of consensus"
repeat_region
22328..22772
/note="L2 repeat: matches 2242. .2710 of consensus"
repeat_region
22768..22848
/note="L2 repeat: matches 2661. .2750 of consensus"
misc_feature
23314..24204
/note="CpG island"
misc_feature
23596..23601
evidence=not_experimental
/note="Weak data."
repeat_region
25664..25783
/note="L2 repeat: matches 2588. .2741 of consensus"
repeat_region
25786..26131
/note="MER33 repeat: matches 1. .324 of consensus"
repeat_region
26858..27183
/note="AluY repeat: matches 1. .301 of consensus"
repeat_region
27208..27334
/note="MIR repeat: matches 81. .221 of consensus"
repeat_region
27673..27833
/note="MIR repeat: matches 7. .175 of consensus"
misc_feature
complement(28112..28516)
/note="match: GSS: Em:AQ212338"

misc_feature	28535..29088	
repeat_region	/note="match: GSS: Em:B51862"	
	28560..28684	
misc_feature	/note="MIR repeat: matches 79. .210 of consensus"	
	28588..28828	
misc_feature	/note="match: GSS: Em:B78941"	
	28588..28841	
repeat_region	/note="match: GSS: Em:B88272"	
	28849..29048	
repeat_region	/note="MER20 repeat: matches 2. .214 of consensus"	
	30148..30232	
misc_feature	/note="12 repeat: matches 2610. .2692 of consensus"	
	complement(31265..31862)	
misc_feature	/note="match: GSS: Em:AQ625926"	
	complement(31333..31862)	
misc_feature	/note="match: GSS: Em:AQ625934"	
	complement(31491..31857)	
repeat_region	/note="match: GSS: Em:AQ131221"	
	31733..31875	
repeat_region	/note="12 repeat: matches 2039. .2197 of consensus"	
	33899..33992	
repeat_region	/note="12 repeat: matches 2578. .2667 of consensus"	
	34111..34437	
repeat_region	/note="MER31A repeat: matches 4. .477 of consensus"	
	35235..35300	
repeat_region	/note="MER5A repeat: matches 121. .189 of consensus"	
	35369..35435	
Query Match 7.3%; Score 242; DB 9; Length 132449;		
Best Local Similarity 49.4%; Pred. No. 7.5e-26;		
Matches 719; Conservative 0; Mismatches 725; Indels 12; Gaps 3;		
QY 997 cattaccagtggcagaagaccgcgcgaagtaccagcgaagaattagatgccactccag	1056	
Db 102384 CAGTAGCAGCGGACGTAATAGTAGTGGTAGTAGCAGCAGCTTGACGTGTCGAGGAG	102443	
QY 1057 aggatgacttcgattagattgttacagaatctccccgaggaagtagaattagattagatg	1116	
Db 102444 GAGTAGCAGCAGCAGCAGTGGCCGTAGCTGTGGCAGGAGGAGTAGCAGCAGCAGCAGT	102503	
QY 1117 aagagcaactgagagaatcaacggaagtggaccacggaagaagaccacccaag	1176	
Db 102504 CCGTAGCTGTGGCAGGAGGAGTAGCAGCAGCAGCTTCCGTAGCTGTGGCAGGAGGAG	102563	
QY 1177 aattagatgcactccagagatggatttcgattagacgaactcagagaagagaaca	1236	
Db 102564 TAGCAGCAGCAGCAGTGGCCG-TAGCTGTGGCAGGAGGAGTAGCAGCAGCAGCTGCA	102622	
QY 1237 gaagaacgttagaggagagaacacagaagaagctgcagaagagaagtatcsgaagaa	1296	
Db 102623 GTAGCT----GTGGCAGGAGTAGCAGCAGCAGCTTGCCAGTAGCTGTGGCAGGAGGAG	102677	
QY 1297 ctccagaaggaagaagattagaggcaactccagagatgatttcgcaattagatgaa	1356	
Db 102678 TACACGATCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAG	102737	
QY 1357 ctacattagaagaacacgaagaactgcagaagagaagaacccgttagaggggagaagaa	1416	
Db 102738 TAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAG	102797	
QY 1417 ccgtagaggagagaacacgcttagagggaagaagctgcagaagagaagaagattag	1476	
Db 102798 CAGCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAG	102857	
QY 1477 aggcaactccagagattgacttccaattagaagaacccatcagagaagagaaggggaag	1536	
Db 102858 CAGGAGTAGCAGCAGCAGCTTGACGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAG	102917	
QY 1537 gagaaggaagggagagaaggaagcgttagtagcagtgccagtagtggtggccgaac	1596	
Db 102918 TTCCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAG	102977	
QY 1597 cggtagaagtagtgcctctgctcagcctgtcaaaccaatggctgcgtcccaacggcagatg	1656	

Db 102978 GAGTAGCAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAGCAGCAGCTGTTGCAGTAGCTG	103037	
QY 1657 aaactttatttcgttgatatttagataaacgattttaacgtatgcagaacattacatcccttg	1716	
Db 103038 TGGCAGGAGGAGTAGCAGCAGCAGCAGCTTCCGTAGCTGTGGCAG-----GAGGATTAG	103091	
QY 1717 agccattatttaaacaaactcctcaaggatcctgatgcagagagagcgtgttaacagtaccat	1776	
Db 103092 CAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAGCATCAGTAGTTGCAGTAGCTG	103151	
QY 1777 caaaggaagcacctgttacaagtaccagtggcagtagggcccgogcaagaagtgtccaacgg	1836	
Db 103152 TGGCAGCAGGAGTAGCAGCAGCAGCAGCTTGCCGTAGCTGTGGCAGGAGGAGTAGCAGCAT	103211	
QY 1837 aagaattgatgcgaactccaagagagcatttcgaattagaaggaactgcagaagactccag	1896	
Db 103212 CAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGG	103271	
QY 1897 aggaaggagaatttagtattagaaggagaagaaaccaacggaagaagagccaaagaag	1956	
Db 103272 CAGGAGTAGCAGCATCAGTAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCAGCAG	1033	
QY 1957 gagagccaacagaaggaagtgccagaagaagaatttagaggcaactccagagagacgatt	2016	
Db 103332 TTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAG	103391	
QY 2017 tcgaattagaagaaccaacagaggaagaagtagaagaacccgttagagggagaagaactg	2076	
Db 103392 CAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGG	103451	
QY 2077 cagaaggagaagaagtgggaagaggtacctgcagaagtagaagaagtgggaagaggtacctg	2136	
Db 103452 CAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAG	103511	
QY 2137 cagaagtagaagaagtgggaagaggtaccagaagaagtagaagaaggtaccgcgagaagtag	2196	
Db 103512 CAGTTGCAGCAGCTGTGGCAGGAGTAGCAGCATCGGCAGTTGCAGCAGCTGTGSCAGGAG	103571	
QY 2197 aagaagtgggaaggtaccagaagaagtgggaaggggtaccagaagaagtgggaagaggtac	2256	
Db 103572 GAGTAGCAGCATCGGCAGTTGCAGCAGCTGTGGCAGGAGGAGTAGCAGCATCGGCAGTTG	103631	
QY 2257 cagaagaagtgggaaggttaccagaagaagtgggaagaagtgggaagaagtagaagaagtag	2316	
Db 103632 CAGCAGCTGTGGCAGGAGGAGTAGCAGCATCGGCAGTTGCAGCAGCTGTGSCAGGAGGAG	103691	
QY 2317 aggtaccagcggtagtagaagtagaagtagaccagcggtagtagaagaagaaggtgccagaag	2376	
Db 103692 TAGCAGCATCGGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAGCATCGGCAGTTGCAGTAG	1037	
QY 2377 aagttagaagaagaagaaggaagaagaaacccagtagagggaagaagatgtattacaattag	2436	
Db 103752 CTGTGGCAGGAGTAGCAGCAGCGGCAGTTGCAGCAGCTGTGGCAGGAGTAGCAGCAGCAG	103811	
QY 2437 taataccatcgggaaga 2452		
Db 103812 CAGTTGCAGTGCCAGGA 103827		
RESULT 28		
AC018348/c		
LOCUS	178022 bp	DNA linear
DEFINITION	Homo sapiens chromosome 15 clone RP11-483E23, WORKING DRAFT	
	SEQUENCE, 9 unordered pieces.	
ACCESSION	AC018348	
VERSION	AC018348.11	GI:12408462
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 (bases 1 to 178022)	

AUTHORS

Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D.,
 Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,
 Mao, J., Lam, B., Marathe, R., Miranda, M., Morehouse, A.J., Nguyen, M.,
 Oefner, P., Palm, C.J., Ramirez, D., Southwick, A.M., Wilhelmy, J.,
 Yu, S. and Davis, R.W.

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 178022)
 Bruno, D., Conn, L., Dela Rosa, M., Federspiel, N., Foreman, P.,
 Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,
 Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J.,
 Yu, S. and Davis, R.W.

TITLE

JOURNAL

Submitted (09-DEC-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA

COMMENT

On Jan 24, 2001 this sequence version replaced gi:12043889.

----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center

Center code: SDSTDC

Web site: <http://sequence-www.stanford.edu/group/human/>

Contact: hum-info@sequence.stanford.edu

----- Project Information

Center project name: 732

Center clone name: RP11-483E23

----- Summary Statistics

Sequencing Vector: M13mp18; X02513

Chemistry: Dye-primer; 11% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 172592 bases at least Q40

Consensus quality: 174564 bases at least Q30

Consensus quality: 175289 bases at least Q20

Insert size: 187151; agarose-fp

Insert size: 177222; sum-of-contigs

Quality coverage: 8.3x in Q20 bases; agarose-fp

Quality coverage: 8.8x in Q20 bases; sum-of-contigs.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 9 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 2877: contig of 2877 bp in length

* 2878 2977: gap of unknown length

* 2978 9281: contig of 6304 bp in length

* 9282 9382: gap of unknown length

* 9382 15085: contig of 5704 bp in length

* 15086 15185: gap of unknown length

* 15186 25650: contig of 10465 bp in length

* 25651 25750: gap of unknown length

* 25751 49398: contig of 23648 bp in length

* 49399 49499: gap of unknown length

* 49499 69237: contig of 19639 bp in length

* 69238 88242: contig of 19005 bp in length

* 88243 88342: gap of unknown length

* 88343 124370: contig of 36028 bp in length

* 124371 124471: gap of unknown length

* 124471 178022: contig of 53552 bp in length.

* Location/Qualifiers

1..178022

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="15"

/clone="RP11-483E23"

/clone_lib="RPC1 human BAC library 11"

1..2877

/note="assembly_name:Contig19"

2978..9281

/note="assembly_name:Contig20"

FEATURES

source

misc_feature 9382..15085
 /note="assembly_name:Contig21"
 misc_feature 15186..25650
 /note="assembly_name:Contig22"
 misc_feature 25751..49398
 /note="assembly_name:Contig23"
 misc_feature 49499..69137
 /note="assembly_name:Contig24"
 clone_end:SP6
 misc_feature 69238..88242
 /note="assembly_name:Contig25"
 clone_end:T7
 misc_feature 88343..124370
 /note="assembly_name:Contig26"
 misc_feature 124471..178022
 /note="assembly_name:Contig27"
 BASE COUNT 51863 a 38471 c 37841 g 49044 t 803 others
 ORIGIN

Query Match 7.3%; Score 242; DB 2; Length 178022;

Best Local Similarity 50.1%; Pred. No. 7.4e-26;

Matches 691; Conservative 0; Mismatches 675; Indels 14; Gaps 3;

QY 1056 gaggatgacttcgattagattgttacagaatctcccgaggaagtagaattagattagat 1115
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149970 GCGGAGCAGGAGGAGATGATGTGGGAGAGAGGAGGAGAGATGCGGAGGAGGAGAT 149911
 QY 1116 gaagaggaactgaagaagaatcaacggaagtgggacccaacggaagaccacacaa 1175
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149910 GATGTGGGAGAGGAGGAGAGATACGGGAGCTGGGAGAGAGAGATGCCAGCAGGAGGAGA 149851
 QY 1176 gaattagatgccactccagaggtgatttcgattagacgaaactcgaaagagagaac 1235
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149850 GATACGGGAGCAGGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 149792
 QY 1236 agaagaac-----gtagagggagaagaacacagaagaagctgcagaagagaatcaga 1291
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149791 GGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 149732
 QY 1292 agaactccagaagagaagaagtagtagaggcaactccagagagatttcgattag 1351
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149731 AGATACGGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 149672
 QY 1352 tggaaactacattagaagaaccgaagaactgcagaagaggaagaacccgtagsggaga 1411
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149671 AGATACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 149612
 QY 1412 agaaccgttagagggagaagaaccgttagaggagaagaagctgcagaagagaagaaga 1471
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149611 TAAGGGAGCAGGAGGAGGAGATATGGAGGCAAAAGGAGAGAGATGCACGAGCAGGAGGAGA 149552
 QY 1472 gttagaggaactccagaggtgacttccattagaagaaccatcagagaagagaag 1531
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149551 TACGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 149492
 QY 1532 ggaagagaaggaaggggaaggaaggaaggaagcgttagtagcagtcagcagtagtggc 1591
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149491 AGATACGGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 149432
 QY 1592 cgaaccggtagaagtagtagtactcctcagcctgtcaaaccaatggtcctccacagc 1651
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149431 AGAAGATGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 149372
 QY 1652 agatgaactttattcgttgatattcttagataaacgatttaacgtatgcacattacatc 1711
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149371 AGAAGATGTAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 149312
 QY 1712 ctttgagccattattaaacaaatcctcaaggatccttgatgcaggagagcgtgttaacagt 1771
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 149311 AGATGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 149252
 QY 1772 accatcaaggaagcaccctgtacaagtaccagtggcagtaggggcccgcgaagaagtgc 1831

RESULT	30
AC068037	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

99063792
2 (bases 1 to 206136)
Isak, A., Kozłowicz, A. and Creason, K.

Waterston, R.H.
Direct Submission
Submitted (27-APR-2000) Genome Sequencing Center, Washington

Waterston, R.H.
Direct Submission
Submitted (03-JAN-2002) Genome Sequencing Center, Washington

Waterston, R.
Direct Submission
Submitted (09-JAN-2002)
University, 4444 Forest Park Avenue
Department of Genetics, Washington
St Louis, Missouri 63108
USA

Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >=

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male

Tateno, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E.,


```
QY 2016 ttccaattagaagaacacacaggaagaagtagaagaacccgtagaggcggaagaact 2075
    ||| ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 77753 GGAGAAGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77812

QY 2076 gcagaaggagaagaagtgggaagggtaccctgcagaagtagaagaagtgggaagggtacct 2135
    | ||||| ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 77813 GGAGAAGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77872

QY 2136 qcagaagttagaagaagtgggaagggtaccagaagaagttagaagaagttaccgcagaagta 2195
    | ||||| ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 77873 GGAGAAGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77932

QY 2196 gaagaagtgggaagggtaccagaagaagtgggaagggtaccagaagaagtgggaagggtta 2255
    | ||||| ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 77933 GGAGAAGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77992

QY 2256 ccagaagaagtgggaagggtaccagaagaagtgggaagaagtgggaagaagttagaagaagta 2315
    ||| ||||| ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 77993 GGAGAAGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 78052

QY 2316 gaggtaccagcggtagtagaagttagaagttaccagcggttagtagaagaaggtgccagaa 2375
    || | | | | ||||| ||||| | ||||| | ||||| | ||||| | |||||
Db 78053 GAAGGAGAGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGA 78112

QY 2376 gaagtagaagaagaagaagggaagaccagtagaggaagaagaatgtattacaa' 2432
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78113 GAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 78169
```

Search completed: June 14, 2002, 17:02:22
Job time: 24633 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 15:29:46 ; Search time 2486.21 Seconds
(without alignments)
18115.662 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gaattccggttaagtaacaa.....tctataataataataattc 3337

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	423	12.7	633	12	AZ569608 263PVD09
2	34	1.0	337	12	AZ852656
3	33	1.0	825	12	AZ682647
4	32	1.0	189	10	BF854267
5	32	1.0	229	10	BF854267
6	32	1.0	250	10	BI708633
7	32	1.0	380	9	AW101380
8	32	1.0	905	12	AZ677174
9	31	0.9	774	10	BG618297
10	30	0.9	144	10	BG959271
11	30	0.9	166	9	BE146466
12	30	0.9	208	10	T05426
13	30	0.9	255	9	AW886018
14	30	0.9	304	9	AW619681
15	30	0.9	335	9	AU097484
16	30	0.9	362	9	AI650676
17	30	0.9	367	10	BM169373

18	30	0.9	383	12	AZ656316
19	30	0.9	396	10	BG683139
20	30	0.9	403	10	BM034050
21	30	0.9	586	10	BI284201
22	30	0.9	627	12	AZ836143
23	30	0.9	785	12	BI150657
24	30	0.9	824	10	BE488177
25	30	0.9	835	12	BI431115
26	30	0.9	1087	12	AQ752154
27	30	0.9	1120	10	BM464445
28	29	0.9	30	12	TA247F06P
29	29	0.9	34	12	TA222B09P
30	29	0.9	68	12	CNS048HA
31	29	0.9	68	12	CNS048HA
32	29	0.9	104	9	AW885174
33	29	0.9	112	9	AL641523
34	29	0.9	113	12	BH500312
35	29	0.9	119	9	AW860980
36	29	0.9	122	10	BE974031
37	29	0.9	133	10	BG995888
38	29	0.9	134	9	AW071377
39	29	0.9	135	9	AW889107
40	29	0.9	138	9	AA790025
41	29	0.9	143	9	AW345955
42	29	0.9	148	10	BF831412
43	29	0.9	149	10	BG981647
44	29	0.9	150	10	BF924884
45	29	0.9	151	10	BG736025

ALIGNMENTS

RESULT 1
AZ569608
LOCUS 633 bp DNA linear GSS 15-MAY-2001
DEFINITION 263PVD09 Pv MBN #30 Plasmodium vivax genomic 3', DNA sequence.
ACCESSION AZ569608
VERSION AZ569608.1 GI:13979872
KEYWORDS GSS.
SOURCE malaria parasite P. vivax.
ORGANISM Plasmodium vivax
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 633)
AUTHORS Carlton,J.N.-R. and Dame,J.B.
TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL Parasitol. Today 16 (10), 409 (2000)
COMMENT Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: M13(-20) forward
Class: shotgun.

FEATURES
Location/Qualifiers
1..633
/organism="Plasmodium vivax"
/strain="Salvador 1 (Collins, W. 1972. J. Parasitol. 69, 497-598)"
/db_xref="taxon:5855"
/clone_lib="pv MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Salmiri boliviensis"
/note="Vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site.1: EcoR V; Site.2: EcoR V; Host leukocytes were extracted from P. vivax infected blood using the following methods: first, infected blood was activated by the addition of 0.5 ml of ADP (40mg/ml) per 10 ml blood. Then blood was passed over a column of acid washed 0.1 mm glass beads, then through a Plasmidpur filter, followed by passage through a column of pre-wet

Whatman CF11 powder (1:2 ratio volume of blood to CF11), and finally centrifuged through a 50% Percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 44% formamide at 50°C as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA Polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."

BASE COUNT 206 a 84 c 196 g 146 t 1 others
ORIGIN

Query Match 12.7%; Score 423; DB 12; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.6e-90;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 gatgtttagagtgaggatcctcaaacgacggttttagaattagaagaggaaatttt 320
|||||
Db 1 GATGTTTACAGGCTGGAGGATCCTTCAACGACGCTTTAGAAATTAGAAGAGGAAAAATTT 60
|||||

QY 321 gatgagaattcaggtgatgaataactcttttagatgctactccccgaagatgactttgcc 380
|||||
Db 61 GATGAGAATTTCAGTGATGATGAATACTCTTTTAGATGCTACTCCCGCAAGATGACTTTGCC 120
|||||

QY 381 ttaacgatttgcgaattgaagcagtgaggaaagtcacaaacgtaacgcttagatgagtgaa 440
|||||
Db 121 TTAACAGATTGCCAATTGAAGCAGATGAGCAAGTCAACGAACGTTAGATGGAGGTGAA 180
|||||

QY 441 tcattgagaggtttccactgaagtatggaacagaagatggctcaacagatgatacg 500
|||||
Db 181 TCATTAGGAGAGGTTTCCACTGAAGATATGGAACAGAGATGGCTCAACAGATGATACG 240
|||||

QY 501 gaacagaacaaaggactacgtgtgatgaaaggaagaaagaaactggcgatatgaa 560
|||||
Db 241 GAACAGAAGCAAGCACTACCTGGTGATATGGAAGGAGAAAGCAAGCTGGCGATATGGAA 300
|||||

QY 561 gcaggggaaagaactggtgtattggaagcagggggaagaaactggcgatttgggaagcagg 620
|||||
Db 301 GCAGGGGAACAAGCTGCTGATTTTGAAGCAGGGAAGAACTGCGCATTTTGAAGCAGGG 360
|||||

QY 621 gaagaaactggcgatttgggaagcagggggaagaaactggcgatttgggaagcagggggaagaa 680
|||||
Db 361 GAAGAACTGGCGATTGGAAGCAGGGAAGAAAGCTGCTGATTGGAAGCAGGGGAAGAA 420
|||||

QY 681 act 683
|||
Db 421 ACT 423

RESULT 2
AZ852656/c 337 bp DNA linear GSS 21-FEB-2001
LOCUS 2M0155K15R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
DEFINITION clone UUGC2M0155K15 R, DNA sequence.
ACCESSION AZ852656
VERSION AZ852656.1 GI:13039991
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 337)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0155 row: K column: 15
Seq primer: CACACAGGAACACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 337.

FEATURES
source

1. .337
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0155K15"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 20 a 140 c 45 g 132 t
ORIGIN

Query Match 1.0%; Score 34; DB 12; Length 337;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1533 gaaggagaaggagaagggaaggagaaggagaag 1566
|||||
Db 306 GAAGGAGAGGAGAGGGGAGGAGGAGGAGGAGAG 273

RESULT 3
AZ852647/c 825 bp DNA linear GSS 14-DEC-2000
LOCUS ENTLI55TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
DEFINITION AZ852647
ACCESSION AZ852647
VERSION AZ852647.1 GI:11819793
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 825)
AUTHORS Loftus, B., Van Aken, S., and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL WHI:IMS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

the library construction is described in

Db 742

Db 742 A

Db 742 AAAATTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT

[illegible]

```
RESULT 10
BG959271/c
LOCUS
DEFINITION CM2-CT0781-190301-701-a01_1 CT0781 Homo sapiens cDNA, mRNA
sequence.
ACCESSION BG959271
VERSION BG959271.1 GI:14377442
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 144)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2at2-CM2-CT0781-
190301-701-a01_1&t3=2001-03-19&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 118.
Location/Qualifiers
1..144
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0781"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 78 a 29 c 23 g 14 t
ORIGIN
Query Match 0.9%; Score 30; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 192 gaaataattttttttttttttttttttttttt 221
|||||
Db 90 GAAAAAATTTTTTTTTTTTTTTTTTTTTT 61
RESULT 11
BE146466/c
LOCUS
DEFINITION QV0-HT0216-011199-043-b06 HT0216 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE146466
VERSION BE146466.1 GI:8609190
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 166)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2-QV0-HT0216-011
199-043-b06&t3=1999-11-01&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 85.
Location/Qualifiers
1..166
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0216"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 61 a 27 c 42 g 36 t
ORIGIN
Query Match 0.9%; Score 30; DB 9; Length 166;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps
QY 194 aaaaattttttttttttttttttttttttt 223
|||||
Db 84 AAAAAATTTTTTTTTTTTTTTTTTTTTTCT 55
RESULT 12
T05426/c
LOCUS
DEFINITION T05426 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
clone HFCW86 similar to EST containing Alu repeat, mRNA sequence.
ACCESSION T05426
VERSION T05426.1 GI:316578
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 208)
Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
3,400 expressed sequence tags identify diversity of transcripts
from human brain
Nature Genet. 4, 256-267 (1993)
9336420
Contact: Adams, MD
```


The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13-21.

FEATURES

source
1. .208
/organism="Homo sapiens"
/db_xref="ATCC (inhost):82144"
/db_xref="taxon:9606"
/clone_lib="HFCW86"
/note="Vector: Fetal brain, Stratagene (cat#936206),
oligo-dT + random primed cDNA synthesis; LambdaZAP-II
vector, 1.0kb average inser size." 3 others

BASE COUNT 71 a 37 c 53 g 44 t
ORIGIN

Query Match 0.9%; Score 30; DB 10; Length 208;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttt 222
|||||
Db 200 AAAAAAATTTTTTTTTTTTTTTTTTTT 171

RESULT 13

AW886018/c
LOCUS
DEFINITION RC1-OT0075-120400-011-h07 OT0075 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW886018
VERSION
KEYWORDS EST.
SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 255)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL
MEDLINE
COMMENT Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=&t2=RC1-OT0075-120
400-011-h07&t3=2000-04-12&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 41

High quality sequence stop: 254.

Location/Qualifiers

FEATURES

source
1. .255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="OT0075"
/dev_stage="Adult"
/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:

SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 102 a 50 c 37 g 66 t
ORIGIN

Query Match 0.9%; Score 30; DB 9; Length 255;
Best Local Similarity 100.0%; Pred. No. 7.2e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 gaaaaattttttttttttttttttttttt 221
|||||
Db 192 GAAAAAATTTTTTTTTTTTTTTTTTTT 163

RESULT 14

AW619681/c
LOCUS
DEFINITION 7754 MARC PBE Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AW619681
VERSION
KEYWORDS EST.
SOURCE pig.

ORGANISM

Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 304)
Smith, T.P.L., Fahrenkrug, S.C., Rohrer, G.A., Simmen, F.A., Rexroad,
C.E. and Keeler, J.W.

Mapping of expressed sequence tags from a porcine early embryonic
cDNA library
Anim. Genet. 32 (2), 66-72 (2001)
21314990

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers

FORWARD: GGAAACAGCTATGACCATG

BACKWARD: GTAAACGACGCCAGT

Seq primer: AATTACCTCTACTAAAGG.

FEATURES

source

1. .304
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC PBE"
/tissue_type="Day 12 whole embryos"
/lab_host="XLOLR"

/note="Vector: pBLUESCRIPT SK-; Site_1: EcoRI; Site_2:
XhoI; Library made from pool of embryos in spherical and
filamentous stages of development (7.5% and 92.5%,
respectively, of each stage) as described in Choi et al,
Endocrinology 137, 1457-67, 1996."

BASE COUNT 107 a 64 c 81 g 52 t
ORIGIN

Query Match 0.9%; Score 30; DB 9; Length 304;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 aaaaaattttttttttttttttttttttt 223
|||||

Db 254 AAAAAAATTTTTTTTTTTTTTTTTTTT 225

RESULT	15
LOCUS	AU097484/c
DEFINITION	Rice shoot Oryza sativa cdna clone S5277, mRNA sequence.
ACCESSION	AU097484
VERSION	AU097484.1 GI:8860166
KEYWORDS	EST.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza. 1 (bases 1 to 335) Sasaki,T. and Yamamoto,K. Rice cDNA from etiolated shoot (2000) Unpublished (2000) Contact: Takuji Sasaki National Institute of Agrobiological Resources Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan Tel: 81-298-38-7441 Fax: 81-298-38-7468 Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/ PROJECT "IRGP"
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

Fax: 01 250 30 7400
 Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/PROJECT/'RGP'

```
S5277_82.. Location/Qualifiers
      1..335
      /organism="Oryza sativa"
      /strain="Nipponbare, sub_species Japonica"
      /db_xref="taxon:4530"
      /clone="S5277"
      /clone_lib="Rice shoot"
      /note="Etiolated shoot (8 days old)"
```

```

ORIGIN
Query Match          0.9%; Score 30; DB 9; Length 335;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY    200   ttttttttttttttttttttttgcctacata 229
      |||||
Db     312   TTTTtttttttttttttttttttttGCTCTACA 283

RESULT 16
AI650676/c
LOCUS
DEFINITION
AI650676 linear EST 17-DEC-1999
wb25d06.xl NCI_CCAP_GC6 Homo sapiens cDNA clone IMAGE:2306699 3'
similar to contains element MER4 repetitive element ; , mRNA
sequence.
ACCESSION AI650676
VERSION AI650676.1 GI:4734655
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 362)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

```
/note=vector: PWD4ZNV;
```

(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT
ORIGIN

Query Match 0.9%; Score 30; DB 10; Length 396;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 193 aaaaaattttttttttttttttttttttttt 222
|||||
Db 88 AAAAAAATTTTTTTTTTTTTTTTTTTT 117

RESULT 20
BM034050/c

LOCUS
DEFINITION
kh72c01.y1 Ascaris suum female gonad DZ pamp1 v2 Chiapelli McCarter
PRECURSOR ; , mRNA sequence.

ACCESSION
BM034050
VERSION
BM034050.1 GI:16747621

KEYWORDS
SOURCE
EST.
pig roundworm.
Ascaris suum

ORGANISM

REFERENCE
1 (bases 1 to 403)
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
; Ascarididae; Ascaris.

AUTHORS

McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
Gibbons, M., Kitter, E., Bennett, J., Franklin, C., Tsagareishvili, R.,
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
Wilson, R.

TITLE
The Washington Univ. Nematode EST Project, 1999

JOURNAL
Unpublished (1999)

COMMENT

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

High quality sequence stop: 115.

FEATURES

source

1..403
Location/Qualifiers
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone_lib="Ascaris suum female gonad DZ pamp1 v2
Chiapelli McCarter"
/sex="Female"
/tissue_type="Dissected female gonad (DZ-differentiation
zone, middle region of gonad)"
/dev_stage="Adult"
/lab_host="DH10b"

/note="Vector: pAMP1 (Gibco); Site.1: Not1; Site.2: SalI;
The library was constructed by Brandi Chiapelli and Dr.
James McCarter at Washington University, St. Louis. The
cDNA was made by using Dynabead oligo-dT priming (Dynal).
PCR based library using a modified protocol from the
SMART PCR cDNA Synthesis Kit from Clontech. Directionally
cloned into the UDG sites of pAMP1. Dissected nematode
tissues were provided by Dr. Alan Scott (ascott@hshp.edu
) of the School of Public Hygiene and Public Health at
John Hopkins University in Baltimore, MD."

BASE COUNT
ORIGIN

227 a 43 c 82 g 51 t

Query Match 0.9%; Score 30; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 192 gaaaaattttttttttttttttttttttttt 221
|||||
Db 178 GAAAAAATTTTTTTTTTTTTTTTTTTT 149

RESULT 21
BI284201/c

LOCUS

DEFINITION
UI-R-CX0s-ccn-d-08-0-UI-s1 UI-R-CX0s Rattus norvegicus cDNA clone

ACCESSION
BI284201

VERSION
BI284201.1 GI:14936667

KEYWORDS
EST.

SOURCE
Norway rat.

ORGANISM
Rattus norvegicus

REFERENCE
1 (bases 1 to 586)

AUTHORS
Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE
Normalization and subtraction: two approaches to facilitate gene

JOURNAL
discovery

MEDLINE
Genome Res. 6 (9), 791-806 (1996)

COMMENT
97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

non-normalized rat placenta pool library cDNA Library Preparation:

M.B. Soares Lab Clone distribution: clones will be available

through Research Genetics (www.resgen.com) the following repetitive

elements were found in this cDNA sequence: 40-176,

>POLY_ASimple_repeat 206-228, >GC_rich#Low_complexity 333-353,

>AT_rich#Low_complexity

Seq primer: M13 Forward

POLYA=yes

Location/Qualifiers

1..386

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-CX0s-ccn-d-08-0-UI"

/dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-CX0s

library is a non-normalized library constructed from the

following rat placenta tissues: embryonic day 17,

embryonic day 19, embryonic day 21. For a detailed

description of the library from which this clone was

derived, please visit our web site at

ratseq.eng.uiowa.edu. The subtraction has been previously

described in (Bonaldo, Lennon and Soares, Genome Research

6:791-806, 1996)

TAG_LTB=UI-R-CX0s

TAG_TISSUE=rat placenta pool

TAG_SEQ=TCACGACAGT"

BASE COUNT 263 a 104 c 143 g 76 t

ORIGIN

Query Match 0.9%; Score 30; DB 10; Length 586;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttg 222
|||||
Db 68 AAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTC 39

RESULT 22

A2836143

LOCUS

DEFINITION A2836143 627 bp DNA linear GSS 20-FEB-2001
clone UUGC2M0130N15 R, DNA sequence.

ACCESSION

A2836143

VERSION

A2836143.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL

COMMENT

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0130 row: N column: 15

Seq primer: CACACAGGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 627.

Location/Qualifiers

1. 627

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0130N15"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F"

/note="Vector: PWD42nv; Purified genomic DNA from M. Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

258 a

82 c

176 g

111 t

ORIGIN

Query Match 0.9%; Score 30; DB 12; Length 627;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1520 agaagagaaagggaaggagaaggagaagg 1549
|||||
Db 15 AGAAGGAGAGGGGAGGAGAGGAGGAGG 44

RESULT 23

BH150657

LOCUS

DEFINITION BH150657 785 bp DNA linear GSS 27-AUG-2001
ENTPK51TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION

BH150657

VERSION

BH150657.1

KEYWORDS

GSS.

SOURCE

Entamoeba histolytica.

ORGANISM

Entamoeba histolytica

REFERENCE

AUTHORS

1 (bases 1 to 785)

Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.

Determination of clone end sequences from Entamoeba histolytica

HMI-IMSS sheared DNA library (2001)

Unpublished (2001)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjoftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI-IMSS sheared DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 17

High quality sequence stop: 325.

Location/Qualifiers

1. 785

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHD51; Site 1: Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 205 a 156 c 170 g 254 t
ORIGIN

Query Match 0.9%; Score 30; DB 12; Length 785;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttg 222
|||||
Db 724 AAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTC 753

COMMENT

Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of *Trypanosoma brucei* (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (to give a tight size distribution (4 kb). The $\nu + i$ method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. M. Vaudin, Academic Press, London, 1997).

Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source
1..34
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="222b09"

BASE COUNT 12 a 0 c 0 g 22 t
ORIGIN

Query Match 0.9%; Score 29; DB 12; Length 34;

Best Local Similarity 100.0%; Pred. No. 9e+03; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttt 221
|||||
Db 6 AAAAAATTTTTTTTTTTTTTTTTTTTTTTT 34

RESULT 30

CNS048HA

LOCUS 68 bp DNA linear GSS 19-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
090014 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION

AL279271.1 GI:8016612

VERSION GSS; genome survey sequence.

KEYWORDS Tetraodon nigroviridis.

SOURCE

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 68)

AUTHORS

Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.

Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

Unpublished

REFERENCE 2 (bases 1 to 68)

AUTHORS

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

Unpublished

REFERENCE 3 (bases 1 to 68)

AUTHORS

Genoscope.

Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source

1..68
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="090014"
/clone_lib="G"
/note="Genoscope sequence ID : COBG090BH07SP1-end :
PUC-Ori"

BASE COUNT 29 a 2 c 1 g 35 t 1 others

Query Match

0.9%; Score 29; DB 12; Length 68;

Best Local Similarity 100.0%; Pred. No. 4.5e+03;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttt 221
|||||

Db 28 AAAAAATTTTTTTTTTTTTTTTTTTTTTTT 56

Search completed: June 14, 2002, 19:14:07

Job time: 13461 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 17:26:02 ; Search time 352.06 Seconds
(without alignments)
16273.770 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gaattccgtaagtaacaa.....tctataaataataatc 3337

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3337	100.0	3337	17 AAT34620	P. vivax ESP-1 blo
2	3337	100.0	3337	20 AAX15174	DNA encoding a sec
3	3337	100.0	3337	22 AAH76457	Plasmodium vivax E
4	33	1.0	6862	22 AAS46299	Tumour suppressor
5	33	1.0	6862	24 ABL32222	Human immune syste
6	33	1.0	6862	24 AAS61081	Human gene regulat
7	30	0.9	334	22 AAI82170	Human polynucleoti
8	30	0.9	337	22 AAL11387	Human breast cance
9	30	0.9	363	22 AAL20286	Human breast cance

c 10	30	0.9	392	22 AAI90039	Human polynucleoti
c 11	30	0.9	5856	23 ABL09674	Drosophila melanog
c 12	30	0.9	8801	23 AAS45437	Chemically pretrea
c 13	30	0.9	8801	24 ABL33741	Human immune syste
c 14	29	0.9	32	13 AAQ43973	Triple helix formi
c 15	29	0.9	40	14 AAQ25032	Oligonucleotide sp
c 16	29	0.9	325	22 AAI89791	Human polynucleoti
c 17	29	0.9	338	22 AAI80023	Human polynucleoti
c 18	29	0.9	344	22 AAI85033	Human polynucleoti
c 19	29	0.9	352	22 AAI91956	Human polynucleoti
c 20	29	0.9	361	22 AAI84186	Human polynucleoti
c 21	29	0.9	369	22 AAI84071	Human polynucleoti
c 22	29	0.9	404	22 AAI82188	Human polynucleoti
c 23	29	0.9	406	22 AAI87091	Human polynucleoti
c 24	29	0.9	416	22 AAI87993	Human polynucleoti
c 25	29	0.9	417	22 AAI84767	Human polynucleoti
c 26	29	0.9	422	22 AAI89825	Human polynucleoti
c 27	29	0.9	425	22 AAI84618	Human polynucleoti
c 28	29	0.9	467	22 AAK63939	Human immune/haema
c 29	29	0.9	608	22 AAH00504	Trichophyton menta
c 30	29	0.9	758	22 AAI88874	Human polynucleoti
c 31	29	0.9	767	22 AAI88875	Human polynucleoti
c 32	29	0.9	821	21 AAO2473	Human colon cancer
c 33	29	0.9	910	24 ABL34307	Human immune syste
c 34	29	0.9	1139	22 AAK86252	Human immune/haema
c 35	29	0.9	1873	21 AAC76766	Human ORFX ORF2321
c 36	29	0.9	2586	22 ABA14588	Human nervous syst
c 37	29	0.9	3270	22 AAS26707	Human genomic DNA
c 38	29	0.9	4978	22 AAK89129	Human digestive sy
c 39	29	0.9	4978	22 AAS31879	Human liver associ
c 40	29	0.9	5690	22 AAS45368	Chemically pretrea
c 41	29	0.9	5752	22 AAF85087	Nucleotide sequenc
c 42	29	0.9	6025	22 AAS45338	Chemically pretrea
c 43	29	0.9	6025	22 AAS45338	Chemically pretrea
c 44	29	0.9	6250	22 AAL36600	Human musculoskele
c 45	29	0.9	6251	22 AAL36603	Human musculoskele

ALIGNMENTS

RESULT 1					
AAT34620					
ID AAT34620	standard; DNA; 3337 BP.				
XX					
AC AAT34620;					
XX					
DT 12-NOV-1996	(first entry)				
XX					
DE P. vivax ESP-1	blood stage antigen coding sequence.				
XX					
KW ESP-1; blood stage	antigen; diagnosis; malaria; infection;				
KW causative agent;	antibody; monoclonal; polyclonal; assay; ds.				
XX					
OS Plasmodium vivax	(clone PvMB3.3.1).				
XX					
FH Key	Location/Qualifiers				
FT Exon	1..91				
FT	/*tag= a				
FT	/note= "encodes initial (N-terminal) sequence of				
FT	hydrophobic amino acids"				
FT	92..230				
FT	/*tag= b				
FT	/note= "contains typical malaria intervening				
FT	sequence splice sites"				
FT	231..3197				
FT	/*tag= c				
FT	exon				
XX					
PN US5532133-A.					
XX					
PD 02-JUL-1996.					
XX					
PF 02-JUN-1993;	93US-0072610.				

Qy	1741	agatcctgctgcagagagcgtgttaacacgtaccatcaaaaggacccctgtacaagtac	1800
Db	1741		1800
Qy	1801	cagtggcagtagggccgcgcaagaagtgcacaaggaagaattgatgcaactccaagagg	1860
Db	1801		1860
Qy	1861	acgatctcgaatttagaaggaaatgcagaagctccagagaaggagaatttagtattagaag	1920
Db	1861		1920
Qy	1921	gagaaggagaaccaacggaagaagccaaagaagaagagaccacaagagaaggaagtgc	1980
Db	1921		1980
Qy	1981	cagaagaagaatttagaggcaactccagagcgatttcggaatttagaagaaccaacagagg	2040
Db	1981		2040
Qy	2041	aagaagtagaagaaaacgtagaggccgaagaacatcgcaagaaggagaagaagtggaaagg	2100
Db	2041		2100
Qy	2101	tacctcagaagttagaagaagtgcgaagaggtacctgcgaagaagtagaagaagtgcgaagagg	2160
Db	2101		2160
Qy	2161	taccagaagaagttagaagaggttaccgcagaagtagaagaagtggaagaggtaccagaag	2220
Db	2161		2220
Qy	2221	aagtgcgaaggttaccagaagaagtgcgaagaaggttaccagaagaagtggaagaggtaccag	2280
Db	2221		2280
Qy	2281	aagaagtgcgaagaagtgcgaagaagtagaagaagtagaggttaccagcggttagtagaagttag	2340
Db	2281		2340
Qy	2341	aagtaccagcggtagtagaagaagaaggtgcgaagaagaagtagaagaagaagaagaagagg	2400
Db	2341		2400
Qy	2401	aagaaccagtagaagaagaagatgtattacaatttagtaataccatcgcgaagaagaataac	2460
Db	2401		2460
Qy	2461	aattagacaaaccaagaagaagcgaattaggctctgcgaattttctctatcatctgcacatgc	2520
Db	2461		2520
Qy	2521	actacaagaagcttccaaaggaaatttatggaagaagaagaagaactgcagtgatccat	2580
Db	2521		2580
Qy	2581	tgaaccagaagaatttgcagaaggaaattcaaatagagtagagcttaataaggcttagagaagaat	2640
Db	2581		2640
Qy	2641	aaggcctagaagcgcactgggaacgattagaagtagcttaataaggcttagagaagaat	2700
Db	2641		2700
Qy	2701	ggaaggacaagaagaataaagaattggcctggcctgcttgcgttaattgaaaaataaaggt	2760
Db	2701		2760
Qy	2761	cagaataatgcataattcaacaaaaggaaaagcccgctgggttttdagaaaaacgagagt	2820
Db	2761		2820

Qy	2821	ggagcgacgagaatggaaaaatggtttaaagcgagagtcacaccccaaatgtattcac	2889
Db	2821	ggagcgacgagaatggaaaaatggtttaaagcgagagtcacaccccaaatgtattcac	2890
Qy	2881	acttgaaaaaatggatgaacgacacatcattccaattttaaataattcttgtaaaata	2940
Db	2881	acttgaaaaaatggatgaacgacacatcattccaattttaaataattcttgtaaaata	2940
Qy	2941	tgtcacaaatttgaaaaacaagaaacccaaagaatgggttaatgaatcactggaaaaagaacg	3000
Db	2941	tgtcacaaatttgaaaaacaagaaacccaaagaatgggttaatgaatcactggaaaaagaacg	3000
Qy	3001	aacgggggttatggtcttgaaatcatttgaagttatgacacacatcaaaattataaatgtgg	3060
Db	3001	aacgggggttatggtcttgaaatcatttgaagttatgacacacatcaaaattataaatgtgg	3060
Qy	3061	ctaagagtcgagaatggtaccgtgcgaatccttaataataatagagaagaagagaactcca	3120
Db	3061	ctaagagtcgagaatggtaccgtgcgaatccttaataataatagagaagaagagaactcca	3120
Qy	3121	tgaatgggttctcttaaaagaaaaacgaatattttagacaaagaatggaaaaaatggact	3180
Db	3121	tgaatgggttctcttaaaagaaaaacgaatattttagacaaagaatggaaaaaatggact	3180
Qy	3181	catctggaaaaagttaaatttttctgttcaattcaatgtgtacaacatttctctggaaaaa	3240
Db	3181	catctggaaaaagttaaatttttctgttcaattcaatgtgtacaacatttctctggaaaaa	3240
Qy	3241	cgccctaaccgaagaatggatcaatcttcttgaatgtaataagaataaagtttgattatgaa	3300
Db	3241	cgccctaaccgaagaatggatcaatcttcttgaatgtaataagaataaagtttgattatgaa	3300
Qy	3301	aaaagaacagattattctcttataataataataattc	3337
Db	3301	aaaagaacagattattctcttataataataataattc	3337
RESULT 2			
AA15174			
ID	AA15174 standard; DNA; 3337 BP.		
XX			
AC	AA15174;		
XX			
DT	28-APR-1999 (first entry)		
DE	DNA encoding a secreted blood-stage protein called PvESP-1.		
KW	Erythrocyte secreted protein-1; PvESP-1; malarial antigen;		
KW	blood-stage protein; malaria; monoclonal antibody 1D1G10; ds.		
XX	Plasmodium vivax.		
OS			
FH	Key	Location/Qualifiers	
FT	CDS	1..3197	
FT		/*tag= a	
FT		/*note= "contains 1 intron"	
FT	exon	1..91	
FT		/*tag= b	
FT		/number= 1	
FT	Intron	92..230	
FT		/*tag= c	
FT		/number= 1	
FT	exon	231..3194	
FT		/*tag= d	
FT		/number= 2	
XX	US5874527-A.		
XX			
PD	23-FEB-1999.		
XX			
PF	30-SEP-1996; 96US-0719822.		
XX			
PR	02-JUN-1993; 93US-0072610.		

PR 07-JUN-1995; 95US-0478417.
 PR 30-SEP-1996; 96US-0719822.
 XX
 PA (UUNY) UNIV NEW YORK STATE.
 XX
 PI Barnwell JW;
 XX
 DR WPI; 1999-180063/15.
 DR P-PSDB; AAW97039.
 XX
 XX Plasmodium vivax peptide antigen - for diagnosis of malaria caused
 PT by Plasmodium vivax
 XX
 PS Example 4; Fig 5A-C; 23pp; English.
 XX
 CC The present sequence encodes a C-terminal erythrocyte secreted
 CC protein-1 (PvESP-1) of Plasmodium vivax. PvESP-1 is a malarial
 CC antigen which is a secreted blood-stage protein present in detectable
 CC amounts in biological samples from individuals infected with P. vivax.
 CC The protein comprises an epitope not present in other Plasmodium species
 CC that cause malaria in humans, and is bound by monoclonal antibody
 CC 1D1G10. The peptide antigen can be used in immunoassays for diagnosis
 CC of malaria caused by P. vivax and/or can be used to produce antibodies
 CC for use in such immunoassays.
 XX
 SQ Sequence 3337 BP; 1304 A; 467 C; 875 G; 691 T; 0 other;

Query Match 100.0%; Score 3337; DB 20; Length 3337;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	gaattccggtaaagttaacactatggttcgtatctatataataacacttaatttacc	60
DB	1	gaattccggtaaagttaacactatggttcgtatctatataataacacttaatttacc	60
QY	61	ttttgttttttttaattcatgcttcaacagtaagataataataatctataaaactgc	120
DB	61	ttttgttttttttaattcatgcttcaacagtaagataataataatctataaaactgc	120
QY	121	tatatatacatatattcataagtgccattgtgaattgcgatacttataatttacgta	180
DB	121	tatatatacatatattcataagtgccattgtgaattgcgatacttataatttacgta	180
QY	181	aaacaatatattgaaaaaatttttttttttttttttttttttttttttttttttttt	240
DB	181	aaacaatatattgaaaaaatttttttttttttttttttttttttttttttttttttt	240
QY	241	aattgaaaaatgcttctgatgtgtgttagagtgagagatccttcaaacgcggtttag	300
DB	241	aattgaaaaatgcttctgatgtgtgttagagtgagagatccttcaaacgcggtttag	300
QY	301	aattagaagaggaaaatttttgatgagaattcaggtgatgataaactcttttagtgcta	360
DB	301	aattagaagaggaaaatttttgatgagaattcaggtgatgataaactcttttagtgcta	360
QY	361	ccccgaagatgactttgcttcaacagatttgccaattgaagacgattgaggaagtcaacg	420
DB	361	ccccgaagatgactttgcttcaacagatttgccaattgaagacgattgaggaagtcaacg	420
QY	421	aaacgcttagatgaggtgaatcattagagaggtttccactgaagatatggaacagagaag	480
DB	421	aaacgcttagatgaggtgaatcattagagaggtttccactgaagatatggaacagagaag	480
QY	481	atggctcaacagatgatacggaacagaagaagactacgtggtgatggaaggagaag	540
DB	481	atggctcaacagatgatacggaacagaagaagactacgtggtgatggaaggagaag	540
QY	541	aagaagctggcgatatgggaagcaggggaagaagctggtgatttgggaagcagggagaaga	600
DB	541	aagaagctggcgatatgggaagcaggggaagaagctggtgatttgggaagcagggagaaga	600
QY	601	ctggcgatttgggaagcaggggaagaactggcgatttgggaagcaggggaagaagctggtg	660

1681	Db	1681	ataacgatttaacgtagtcagacattacatcctttggccattattttaacaaatcctca	1740
1741	Qy	1741	aggatcctgatgcagagagcgtgtaaacagtaccatacaaaaggaagcaccgtgtacaagtac	1800
	Db	1741	aggatcctgatgcagagagcgtgtaaacagtaccatacaaaaggaagcaccgtgtacaagtac	1800
1801	Qy	1801	caagtgcagtagggccgcgcgaagaagtgtccaaacggaagaattgtagtgcactccaagagg	1860
1801	Db	1801	caagtgcagtagggccgcgcgaagaagtgtccaaacggaagaattgtagtgcactccaagagg	1860
1861	Qy	1861	acgatttcgaattagaagaacctgcagaagctccagagctccagaggaagagataattagattagaag	1920
1861	Db	1861	acgatttcgaattagaagaacctgcagaagctccagaggaagagataattagattagaag	1920
1921	Qy	1921	gagaaggaaccaaacggaagaagccaagagaaggaaggaagcaccacaagaaggaagtgc	1980
1921	Db	1921	gagaaggaaccaaacggaagaagccaagagaaggaaggaagcaccacaagaaggaagtgc	1980
2040	Qy	1981	cagaagaagaattagaggcaactccagaggacgatttcgaattagaagaaccacaggag	2040
2040	Db	1981	cagaagaagaattagaggcaactccagaggacgatttcgaattagaagaaccacaggag	2040
2100	Qy	2041	aagaactagaagaacaaacctagaggcggaagaacctgcagaaggaagaagaagtggaaagagg	2100
2100	Db	2041	aagaactagaagaacaaacctagaggcggaagaacctgcagaaggaagaagaagtggaaagagg	2100
2160	Qy	2101	tacctgcagaagttagaagaagtgcgaagaggttacctgcgaagaagttagaagaagtgcgaagagg	2160
2160	Db	2101	tacctgcagaagttagaagaagtgcgaagaggttacctgcgaagaagttagaagaagtgcgaagagg	2160
2220	Qy	2161	taccagaagaagttagaagaaggttaccgcgagaagttagaagaagtggaaagaggtaccagaag	2220
2220	Db	2161	taccagaagaagttagaagaaggttaccgcgagaagttagaagaagtggaaagaggtaccagaag	2220
2280	Qy	2221	aagtgcgaagaggtaccagaagaaggtgcgaagaggttaccagaagaaggtgcgaagaggtaccag	2280
2280	Db	2221	aagtgcgaagaggtaccagaagaaggtgcgaagaggttaccagaagaaggtgcgaagaggtaccag	2280
2340	Qy	2281	aagaagtgcgaagaagtgcgaagaagttagaagaagttagaggttaccagcggtagtagaagaagt	2340
2340	Db	2281	aagaagtgcgaagaagtgcgaagaagttagaagaagttagaggttaccagcggtagtagaagaagt	2340
2400	Qy	2341	aagtaccagcggtagtagaagaagaaggtgcgcagaagaagttagaagaagaagaagaagagg	2400
2400	Db	2341	aagtaccagcggtagtagaagaagaaggtgcgcagaagaagttagaagaagaagaagaagagg	2400
2460	Qy	2401	aagaaccagtagagaagaagaagtgtattacaattagtaataaccatcgcgaagaagaataac	2460
2460	Db	2401	aagaaccagtagagaagaagaagtgtattacaattagtaataaccatcgcgaagaagaataac	2460
2520	Qy	2461	aatagaacaaacaaagaaagacgaatttaggcctctggaaattttctctcatcatcgacatgc	2520
2520	Db	2461	aatagaacaaacaaagaaagacgaatttaggcctctggaaattttctctcatcatcgacatgc	2520
2580	Qy	2521	actaccaaagacgttccaaaggaaattttaaggaaagaagaagaacatgcagttgatccat	2580
2580	Db	2521	actaccaaagacgttccaaaggaaattttaaggaaagaagaagaacatgcagttgatccat	2580
2640	Qy	2581	tgaacccagaagaattttgcgaaggaagattcacaaatctacagaataagctcacattcatc	2640
2640	Db	2581	tgaacccagaagaattttgcgaaggaagattcacaaatctacagaataagctcacattcatc	2640
2700	Qy	2641	aaggcctagaagcgactgggaacgatttagaagtgagcttaaaataaggcttagagaagaagt	2700
2700	Db	2641	aaggcctagaagcgactgggaacgatttagaagtgagcttaaaataaggcttagagaagaagt	2700
2760	Qy	2701	ggatgcgaacaaagaaataaagaattgggctggctggctggcttaattgaaaaataaattggt	2760
2760	Db	2701	ggatgcgaacaaagaaataaagaattgggctggctggctggcttaattgaaaaataaattggt	2760
2820	Qy	2761	cagaataatgtcaaatctcaacaaagaagaagggacccagctggcttttgaaaaaacagagat	2820
2820	Db	2761	cagaataatgtcaaatctcaacaaagaagaagggacccagctggcttttgaaaaaacagagat	2820

Qy	2821	ggagcgcagaaatgggaaaaatgggtttaaagcagaagtcaaatcccaatgtattcac	2880
Db	2821	ggagcgcagaaatgggaaaaatgggtttaaagcagaagtcaaatcccaatgtattcac	2880
Qy	2881	acttgaaaaaatggatgaacgacactcattccaatttatttaaattctctgtgaagata	2940
Db	2881	acttgaaaaaatggatgaacgacactcattccaatttatttaaattctctgtgaagata	2940
Qy	2941	tgtccaaatttgaaaaacaagaaaaaaccaagaatagttaattgaatcactggaagaaagacg	3000
Db	2941	tgtccaaatttgaaaaacaagaaaaaaccaagaatagttaattgaatcactggaagaaagacg	3000
Qy	3001	aacgggggttatggtctctgaaatcatttgaagtattgaccacatcaaaatttataaattcgg	3060
Db	3001	aacgggggttatggtctctgaaatcatttgaagtattgaccacatcaaaatttataaattcgg	3060
Qy	3061	ctaagagtcgagaattggtaccctgccaatccctaataataatagagaagaagagaactca	3120
Db	3061	ctaagagtcgagaattggtaccctgccaatccctaataataatagagaagaagagaactca	3120
Qy	3121	tgaattggtttctctcaagaagaaaacgaattatttagcacaagaattggaagaaatggact	3180
Db	3121	tgaattggtttctctcaagaagaaaacgaattatttagcacaagaattggaagaaatggact	3180
Qy	3181	catttgaaaaaagttaaaatttttggttccaattcgaatggtgtcaacaactttctcgaaaa	3240
Db	3181	catttgaaaaaagttaaaatttttggttccaattcgaatggtgtcaacaactttctcgaaaa	3240
Qy	3241	cgccaaaccaagaagaatggaaatcaattgtttaatgaataaaaagtttgaaattatgaa	3300
Db	3241	cgccaaaccaagaagaatggaaatcaattgtttaatgaataaaaagtttgaaattatgaa	3300
Qy	3301	aaaagaacagatttatctctctataaataaaaataattc	3337
Db	3301	aaaagaacagatttatctctctataaataaaaataattc	3337

RESULT	3
AAH76457	
ID	AAH76457 standard; DNA; 3337 BP.
XX	
XX	AAH76457;
XX	
XX	22-OCT-2001 (first entry)
DE	Plasmodium vivax ESP-1 DNA.
XX	
KW	Plasmodium vivax; ESP-1; erythrocyte secreted protein-1; PvESP-1;
KW	species-specific; malarial peptide antigen; infection; diagnosis;
KW	malaria; ds.
XX	
OS	Plasmodium vivax.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..3197
FT	/tag= a
FT	/product= "ESP-1"
FT	1..91
FT	/tag= b
FT	/number= 1
FT	92..230
FT	/tag= c
FT	/number= 1
FT	231..3197
FT	/tag= d
FT	/number= 2
XX	
PV	US6231861-B1.
XX	
PD	15-MAY-2001.
XX	
PF	05-JUN-1998;
XX	980S-0092458.

Db	1621	agcctgtcaaaccaatggtcgctgccacggcagatgaaactttattcttgcgttgatattcttag	1680
Qy	1681	ataacgatttaacgtatgcagacattacatctcttgagccattatttaacaaactctca	1740
Db	1681	ataacgatttaacgtatgcagacattacatctcttgagccattatttaacaaactctca	1740
Qy	1741	aggatcctgatgcaggagaagcgtgtaacagttaccatcaaggagaacccctgtacaagttac	1800
Db	1741	aggatcctgatgcaggagaagcgtgtaacagttaccatcaaggagaacccctgtacaagttac	1800
Qy	1801	cagtggcagtgaggcccgccgcaagaagtgccaaaggagaattgatgcaactccaagagg	1860
Db	1801	cagtgscagtaggcccgccgcaagaagtgccaaaggagaattgatgcaactccaagagg	1860
Qy	1861	acgatttcgaattagaaggaactcgcaagaagctccagagggaggagaattagttattagaag	1920
Db	1861	acgatttcgaattagaaggaactcgcaagaagctccagagggaggagaattagttattagaag	1920
Qy	1921	gagaaggaaccaacggaaggaagccaagagaaggaagcaccacagaagagagaagtgc	1980
Db	1921	gagaaggaaccaacggaaggaagccaagagaaggaagcaccacagaagagagaagtgc	1980
Qy	1981	cagaagaagaaattagaggcaactccagaggacgatttcgaattagaagaaccacacaggag	2040
Db	1981	cagaagaagaaattagaggcaactccagaggacgatttcgaattagaagaaccacacaggag	2040
Qy	2041	aagaagtacagaacaaccgtagaggcggaagaactgcagaaggagaagaagtggagaagg	2100
Db	2041	aagaagtacagaacaaccgtagaggcggaagaactgcagaaggagaagaagtggagaagg	2100
Qy	2101	tacctcgagaagttagaagaagtgggaagaggttacctgcagaagttagaagaagtggagaagg	2160
Db	2101	tacctcgagaagttagaagaagtgggaagaggttacctgcagaagttagaagaagtggagaagg	2160
Qy	2161	taccagaagaagttagaagaaggttaccgcgagaagtagaagaagtgggaagaggttaccagaag	2220
Db	2161	taccagaagaagttagaagaaggttaccgcgagaagtagaagaagtgggaagaggttaccagaag	2220
Qy	2221	aegtggagaaggttaccagaagaagtgggaagaggttaccagaagaagtggaagaggttaccag	2280
Db	2221	aegtggagaaggttaccagaagaagtgggaagaggttaccagaagaagtggaagaggttaccag	2280
Qy	2281	aagaadtggagaagaagtgggaagaagttagaagaagttagagggttaccagcggtagtagaagaagt	2340
Db	2281	aagaadtggagaagaagtgggaagaagttagaagaagttagagggttaccagcggtagtagaagaagt	2340
Qy	2341	aegtaccagcggtagtagaagaagaaggtggccagaagaagttagaagaagaagaagaagagg	2400
Db	2341	aegtaccagcggtagtagaagaagaaggtggccagaagaagttagaagaagaagaagaagagg	2400
Qy	2401	aagaaccagtagagaagaagaagtgtattacaattagtaataccatcgcggaagaagatatatc	2460
Db	2401	aagaaccagtagagaagaagaagtgtattacaattagtaataccatcgcggaagaagatatatc	2460
Qy	2461	aattagacaaaccaagaagaagcgaattagcctctggaattttatctatcatcgacaatgc	2520
Db	2461	aattagacaaaccaagaagaagcgaattagcctctggaattttatctatcatcgacaatgc	2520
Qy	2521	actaccaaagcgttccaaagggaatttatggagaagaagaagaacgtgcagtgtagtccat	2580
Db	2521	actaccaaagcgttccaaagggaatttatggagaagaagaagaacgtgcagtgtagtccat	2580
Qy	2581	tgaaccagaagaattttgcaagggaagattcacaaattcacaagaattgctcatcatcttc	2640
Db	2581	tgaaccagaagaattttgcaagggaagattcacaaattcacaagaattgctcatcatcttc	2640
Qy	2641	aaggccttagaaggcagctgggaacgatttagaagttagcttaataaaggcctagagaagaat	2700
Db	2641	aaggccttagaaggcagctgggaacgatttagaagttagcttaataaaggcctagagaagaat	2700
Qy	2701	ggatggacaagaagaataaagggaattggcgtggcgtggcgttgcgttaattgaaaaataatgggt	2760
Db	2701	ggatggacaagaagaataaagggaattggcgtggcgtggcgttgcgttaattgaaaaataatgggt	2760

RESULT 4

RESULI
AAS46299

AAS40233
ID AAS46299 standard; DNA: 6862 BP.

AA AAS46299;

18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #21.

Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cytotoxic

KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;

KW cytosine methylation; ds.

OS Homo sapiens.

PN WO200168912-A2.

PD 20-SEP-2001.

15-MAR-2001: 2001WO-EP02955.

XX
PR 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX
XX

XX
CONFIDENTIAL (OFFICIAL)

PI

RESULT 11
ABL09674/c
ID ABL09674 standard; cDNA; 5856 BP.
XX
XX ABL09674;
AC
XX
XX
DT 26-MAR-2002 (first entry)
XX
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 23504.
XX
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-658660/75.
DR
XX P-PSDB; ABB65571.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
XX Claim 1; SEQ ID NO 23504; 2lpp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 5856 BP; 1527 A; 1429 C; 1445 G; 1455 T; 0 other;
SQ

Query Match 0.9%; Score 30; DB 23; Length 5856;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 195 aaaaatttttttttttttttttttttttttttttt 224
|||||
Db 3372 AAAAAATTTTTTTTTTTTTTTTTTTTGTGT 3343

RESULT 12
AAS45437
ID AAS45437 standard; DNA; 8801 BP.
XX
XX
AC AAS45437;
XX
XX
DT 18-DEC-2001 (first entry)
XX
XX
DE Chemically pretreated complementary DNA associated with cell cycle #71.
XX
XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW

human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
XX Homo sapiens.
OS
XX
XX WO200168911-A2.
PN
XX
XX 20-SEP-2001.
PD
XX
XX 15-MAR-2001; 2001WO-EP02945.
PF
XX
XX 15-MAR-2000; 2000DE-1013847.
PR
XX 06-APR-2000; 2000DE-1019058.
PR
XX 07-APR-2000; 2000DE-1019173.
PR
XX 30-JUN-2000; 2000DE-1032529.
PR
XX 01-SEP-2000; 2000DE-1043826.
PR
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX
XX WPI; 2001-602751/68.
DR
XX
XX Designing primers and probes for analysing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle -
XX
XX
XX Claim 1; SEQ ID No 142; 28pp; English.
PS
XX
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers.
XX
XX
XX Sequence 8801 BP; 2344 A; 366 C; 2057 G; 4034 T; 0 other;
SQ

Query Match 0.9%; Score 30; DB 22; Length 8801;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 192 gaaaaatttttttttttttttttttttttttttttt 221
|||||
Db 971 gaaaaatttttttttttttttttttttttttttttt 1000

RESULT 13
ABL33741
ID ABL33741 standard; DNA; 8801 BP.
XX
XX
AC ABL33741;
XX
XX
DT 26-MAR-2002 (first entry)
XX
XX
DE Human immune system associated gene SEQ ID NO: 1714.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;


```
Db 337 AAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTT 309

RESULT 18
AAI85033/c
ID AAI85033 standard; cDNA; 344 BP.
XX
AC AAI85033;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5093.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR P-PSDB; AAO05102.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 5093; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 344 BP; 165 A; 55 C; 74 G; 50 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 344;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttt 221
DB 158 AAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTT 130

RESULT 19
AAI91956/c
ID AAI91956 standard; cDNA; 352 BP.
XX

Db 337 AAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTT 309

RESULT 18
AAI85033/c
ID AAI85033 standard; cDNA; 344 BP.
XX
AC AAI85033;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5093.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR P-PSDB; AAO05102.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 5093; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 344 BP; 165 A; 55 C; 74 G; 50 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 344;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttt 221
DB 158 AAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTT 130

RESULT 19
AAI91956/c
ID AAI91956 standard; cDNA; 352 BP.
XX

AC AAI91956;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 12016.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR P-PSDB; AAO12025.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 12016; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 352 BP; 132 A; 61 C; 66 G; 93 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 352;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttt 221
DB 89 AAAAAAATTTTTTTTTTTTTTTTTTTTTTTT 61

RESULT 20
AAI84186/c
ID AAI84186 standard; cDNA; 361 BP.
XX
AC AAI84186;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4246.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
```



```
XX (HYSE-) HYSEQ INC.
XX PA
XX PI Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AAO02257.
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX Claim 1; SEQ ID NO 2248; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 404 BP; 133 A; 74 C; 83 G; 114 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 404;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 aaaaaattttttttttttttttttttttttt 222
   |||||||
DB 195 AAAAAAATTTTTTTTTTTTTTTTTTTT 167

RESULT 23
AAI87091/c
ID AAI87091 standard; cDNA; 406 BP.
XX
XX AAI87091;
XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 7151.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AAO07160.
```

```
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX Claim 1; SEQ ID NO 7151; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 406 BP; 127 A; 115 C; 66 G; 98 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 406;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttt 221
   |||||||
DB 287 AAAAAAATTTTTTTTTTTTTTTTTTTT 259

RESULT 24
AAI87993/c
ID AAI87993 standard; cDNA; 416 BP.
XX
XX AAI87993;
XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 8053.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AAO08062.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX Claim 1; SEQ ID NO 8053; 1399pp + Sequence Listing; English.
```

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 416 BP; 208 A; 33 C; 96 G; 62 T; 17 other;

Query Match 0.9%; Score 29; DB 22; Length 416;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 aaaaaattttttttttttttttttttttttt 221
|||||
Db 271 AAAAAATTTTTTTTTTTTTTTTTTTTTT 243

RESULT 25
AAI84767/C
ID AAI84767 standard; CDNA; 417 BP.
XX
AC AAI84767;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4827.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO04836.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 4827; 1399pp + Sequence Listing; English.
XX

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 417 BP; 149 A; 80 C; 95 G; 93 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 417;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 aaaaaattttttttttttttttttttttttt 221
|||||
Db 127 AAAAAATTTTTTTTTTTTTTTTTTTTTT 99

RESULT 26
AAI89825/C
ID AAI89825 standard; CDNA; 422 BP.
XX
AC AAI89825;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 9885.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO09894.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 9885; 1399pp + Sequence Listing; English.
XX

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.


```

DT 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 8934.
XX
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX OS Homo sapiens.
XX WO2001I64835-A2.
XX PN
XX XX
XX PD 07-SEP-2001.
XX
XX PF 26-FEB-2001; 2001WO-US04927.
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX PI WPI; 2001-514838/56.
XX DR P-PSDB; AAO08943.
XX PT Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX PS Claim 1; SEQ ID NO 8934; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 758 BP; 257 A; 157 C; 138 G; 204 T; 2 other;

Query Match 0.9%; Score 29; DB 22; Length 758;
Best Local Similarity 100.0%; Pred.No.0.23;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 aaaaaattttttttttttttttttttttt 221
| | | | | | | | | | | | | | | | | |
Db 270 AAAAAAATTTTTTTTTTTTTTTTTTTT 242

Search completed: June 14, 2002, 20:35:46
Job time: 11384 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 14, 2002, 15:37:31 ; Search time 4065.53 Seconds
(without alignments)
17176.568 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gattccggtaaagtaacaa.....tcttataaaataaataattc 3337

Scoring table: OLIGO_NUC
Gapop 60.0 , Capext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

1	3337	100.0	3337	6	ARI51085
2	3337	100.0	3337	6	I23337
3	40	1.2	70206	2	AC067910
4	40	1.2	165868	2	AC027625
5	34	1.0	186562	2	AL451079
6	34	1.0	200898	2	AC096541
7	34	1.0	225609	2	AL645546
8	33	1.0	6862	6	AX251053
9	33	1.0	6862	6	AX251776
10	33	1.0	6862	6	AX345124
11	33	1.0	240000	9	HS421735
12	33	1.0	246176	2	AC096301
13	32	1.0	224010	2	AP001848
14	32	1.0	318221	2	PFMAL13P3
15	31	0.9	35836	3	CEZC334
16	31	0.9	140798	2	AC027803
17	31	0.9	172896	9	AL158193
18	31	0.9	173817	2	AC012141
19	31	0.9	198820	2	AC023315
20	31	0.9	198932	2	AL671090
21	31	0.9	243675	10	AC079869
22	30	0.9	8801	6	AX277981
23	30	0.9	8801	6	AX323672
24	30	0.9	8801	6	AX346643
25	30	0.9	16689	9	HS4011712
26	30	0.9	49816	9	AC093149
27	30	0.9	58569	2	AC015033
28	30	0.9	64255	2	AC024343
29	30	0.9	67462	2	AC091595
30	30	0.9	73858	2	AC109316
31	30	0.9	82419	9	AC004979
32	30	0.9	89203	2	AC010691
33	30	0.9	92491	2	AL672088
34	30	0.9	96874	9	AL138709
35	30	0.9	96879	9	HS384F21
36	30	0.9	98935	9	AC004935
37	30	0.9	99187	9	HS35M08
38	30	0.9	103574	9	HSAC002115
39	30	0.9	113880	3	PFMAL3P4
40	30	0.9	115027	2	AP001480
41	30	0.9	117309	2	AC099091
42	30	0.9	123189	9	AC002556
43	30	0.9	125722	2	AC015695
44	30	0.9	138800	30	AC016215
45	30	0.9	140955	2	AP001971

ALIGNMENTS

RESULT	1	ARI51085	Sequence 1 from patent US 6231861.	3337 bp	DNA	linear	PAT 08-AUG-2001
LOCUS	ARI51085	Sequence 1 from patent US 6231861.					
DEFINITION	ARI51085	Sequence 1 from patent US 6231861.					
ACCESSION	ARI51085	Sequence 1 from patent US 6231861.					
VERSION	ARI51085.1	GI:15117135					
KEYWORDS	Unknown.						
SOURCE	Unknown.						
ORGANISM	Unclassified.						
REFERENCE	1 (bases 1 to 3337)						
AUTHORS	Barnwell,J.W.						
TITLE	Plasmodium vivax blood stage antigens, antibodies, and diagnostic assays						
JOURNAL	Patent: US 6231861-A 1 15-MAY-2001;						
FEATURES	Location/Qualifiers						
source	1..3337						
	/organism="unknown"						
BASE COUNT	1304 a 467 c 875 g 691 t						
ORIGIN							

Query Match

100.0%; Score 3337; DB 6; Length 3337;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	gaattccggtaaaagtaacaactatagtttcgtatctatatataaacttactaattttatc	60
Db	1	GAATTCGGGTAAGTAACAACATATGGTTTCGTATCTATATATAACCTTACTAAATTTTATC	60
Qy	61	ttttgctttctttaaattcattcgtctcaacagtaagataaaaaataatctataaaaactgc	120
Db	61	TTTTGCTTTCTTTAAATTCATGCTTCAACAGTAGATAAAAATAATCTATAAAAACATGC	120
Qy	121	tatatatacatatatttcataatggtggtcatttgtaattgogatcatttaatttcacgta	180
Db	121	TATATATACATATATATTCATAAATGCGCATTTGTCAATTTGCCATCATTTAAATTTACGTA	180
Qy	181	aaaaacaattatgaaaaaatt	240
Db	181	AAAACAATATTGAAAAAATTTTTT	240
Qy	241	aattggaaaatgcttctgatgatgtgttagaggtgagagatccttcaaacgacggttttag	300
Db	241	AATTGGAAAAATGCTTCTGATGATGTTGTAGAGTGTGAGAGTGAGATCCTTCAAAACGACGCTTTAG	300
Qy	301	aattagaagagaaaattttgatgagaattcaggttgatgataaactcttttagatgcta	360
Db	301	AATTAGAAGAGGAAAAATTTTGATGAGAAATTCAGGTGATGATGAAACTCTTTTAGATGCTA	360
Qy	361	cccccgaaagtagcttgccttaacagatttgcctaattgaagacgatgagaaagtcacg	420
Db	361	CCCCCGAAGATGACATTTGCCTTAACAGATTTGCCAANTTGAAGACGATGAGNAAGTCAACG	420
Qy	421	aaacttagatggaggtgaattcattagagaggtttccactgaagatgatgaaacagaag	480
Db	421	AAACGTTAGATGGAGTGAATCATTTAGGAGAGTTCACATGAAGATATGGAACAGAG	480
Qy	481	atggctcaacagatatacaggaacacagaaagagactacctggtgatgaaagagagag	540
Db	481	ATGGCTCAACAGATATACGGAACAGAAAGAGAGACTACCTGGTGATATGGAAGAGAGAG	540
Qy	541	aagaagctggcgatataggaagcaggggaagaagctgggtgatttgaagcagsgggaagaa	600
Db	541	AAGAAGCTGGCCATATGAAGCAGGGGAAGAAGCTGGTGATTGGAACGAGGGGAAGAA	600
Qy	601	ctggcgatttgaagcaggggaagaactggcgatttggaaagcaggggaagaagctggtg	660
Db	601	CTGGCGATTTTGAAGCAGGGGAAGAACTGGCGATTTTGGAAGCAGGGGAAGAAGCTGGTG	660
Qy	661	atttgaagcaggggaagaactggcgatttggaaagcaggggaagaagaactggagatcg	720
Db	661	ATTTGGAAGCAGGGGAAGAACTGGCGATTTCGAAGCAGGGGAAGAACTGGAGATGCGG	720
Qy	721	aaactgaagaagaggaactggagatgcggaaactgaaatggagcaactgtgtatgtag	780
Db	721	AAACTGAAGAAGGAGCAACTGCAGATGCGGAACCTGAAATGGAGCAACTGTGTATGTAG	780
Qy	781	acacagaagatagttcaagctgatgagcagaaaaagtagacatgttctgctcaagaataag	840
Db	781	ACACAGAAGATAGTTCAAGCTGATGAGCAGAGAAAAAGTAGTACATGTTCTGCTCAAGAAAAATG	840
Qy	841	tacaactcgcgatagtaagatgcctcttttggagttatttggataaagataataatt	900
Db	841	TACAACCTGCCGATAGTAATGATGCCCTCTTTGGAAAGTATTTTGGATTAAGATATAATTT	900
Qy	901	ttgatcatattaaaatttcagagccactattcgaacaaattgtgctgggttactgctaacc	960
Db	901	TTGATCATATTAAATTTTCGAGCCACTATTTCGAACAAATTTGTGCGGGTACTGCTAAAC	960
Qy	961	atggttacgggacaagaattgccaatgaacacctgtaccattaccagtgcagaagagcccg	1020
Db	961	ATGTTACGGGACAAGAAATTTGCAATGAACCTGTACCATTTACAGTTGCGCAGAAAGACCCG	1020
Qy	1021	cgaagtagcagcgaagaatttagatgccactccagagagtagacttcgattagatgta	1080
Db	1021	CGAAGTAGCAGCAGAAATTTAGATGCCACTCCAGAGAGTAGACTTCGATTAGATGTTA	1080

Db	1021	CCCAAGTACCAGCGGAGAAATTAGATGCCACTCCAGAGGATGACTTCGCATTAGATGTTA	1080
Qy	1081	cagaatctcccgagaagaatagaatttagtattagatgaagagcgaactgaagaagaatcaa	1140
Db	1081	CAGAATCTCCCGAGGAAGTAGAATTTAGTATTAGTGAAGGCAACTGAAGAAGARTCAA	1140
Qy	1141	cggaaagtggaccacggaagacccaaccgcaagaattagatgccactccagaggatg	1200
Db	1141	CGGAAGTGGACCAACGCGAAGAGCACCAACCGAAGAAATTAGTCCACTCCAGAGGATG	1200
Qy	1201	gatttcgcattagacgaactgcagaagggagaaacagagaacacgtagagggagaagaaa	1260
Db	1201	GATTTCGATTAGACGAAACTGCAGAAGGAGAAACAGAAAGTATAGAGGGAGAGAAA	1260
Qy	1261	cagaagaagctgcagaagagaagtatcagaagaactccagaagagagaagaagattag	1320
Db	1261	CAGAAGAAGCTGCAGAGGAGAAAGTATCAGAAGAAACTTCCAGAGGAGAAAGAGATTAG	1320
Qy	1321	aggcaactccagagagatatttcgcatttagatggaactacattagaagaacccgaagaaa	1380
Db	1321	AGGCAACTCCAGAGGATGATTTCGCATTAGATGGAACTACATTAGAAGAAACCGAAGAAA	1380
Qy	1381	ctgcagaaggaagaacccgtagagggagaagaacccgtagagggagaagaacccgtag	1440
Db	1381	CTGCAGAAGGAGAAACCCGTAGAGGGAGAGAAACCGTAGAGGGAGAAACCCGTAG	1440
Qy	1441	agggaagaagctgcagaagagagaagagattagaggaactccagagagatgacttcc	1500
Db	1441	AGGGAGAAGAAGCTGCACAGAGAGAGAAGAGTTAGAGGCAACTCCAGAGGATGACTTCC	1500
Qy	1501	aattagaagaacctcaggaagagagaaggggaagggagaagggagaagagaag	1560
Db	1501	AATTAGAAGAACCATTAGGAGAAGAGAGGGGAAGGAGAGAGGAGGAGGAGAGAGAG	1560
Qy	1561	gagaagccttagtgcagctgcccagtagtggccgaacccgtagaagtgcactcctctc	1620
Db	1561	GAGAAGCCTTTAGTAGCAGTGCAGTAGTGGCGAACCCTGAGAGTAGTAGTACTCTCTCTC	1620
Qy	1621	agcctgtcaaaccaatgctcctccacggcagatgaaactttattccttgcattcttag	1680
Db	1621	AGCCTGTCAAAACCAATGCTCGCTCCAACGGCAGATGAACCTTATTCGTTGATATCTTAG	1680
Qy	1681	ataacgatttaacgtatgcagacattacatcctttgagccattattaaacaactccca	1740
Db	1681	ATAACGATTTAACGTTATGCAGACATFATACCTCTTTGAGCCATTATTAAACAATCTCTCA	1740
Qy	1741	aggatccctgtagcagagagagcgtgtaacagtagccatcaaaaggaagcactgtacaagtac	1800
Db	1741	AGGATCCTGTAGCAGAGAGGCTGTAACAGTACCATCAAAAGGAGCACCTGTACAAGTAC	1800
Qy	1801	cagtggcagtagggcccgcgaagaagtgcacaacgggaagaaattgatgccaaagagg	1860
Db	1801	CAGTGGCAGTAGGGCCCGCGCAAGAAGTCCCAACGGAAGAAATTGATGCAACTTCCAAGAGG	1860
Qy	1861	acgatttcgaattagaagaactgcagaagctccagagaagagaataattagattagaag	1920
Db	1861	ACGATTTCGAATTAGAAGAACTGCAGAAAGCTCCAGAGAGGAGAAATTAGTATTAGAAG	1920
Qy	1921	gagaagggagaacccaacggaagaagcgaagaagggagagccacagaagaagagaagtgc	1980
Db	1921	GAGAAGGAGAACCAACGGAAGAAGAGCCCAAGAGAGGAGAGCCACAGAAAGGAGAAGTGC	1980
Qy	1981	cagaagaagaatttagaggcaactccagaggacgttccgaattagaagaacccaacagag	2040
Db	1981	CAGAAGAAGAATTAGAGGCACTCCAGAGGACGATTTCGAATTAGAAGAAACCAACAGGAG	2040
Qy	2041	aagaagtagaagaacccgtagagggcgaagaactgcagaaggaagaagaagtggaagag	2100
Db	2041	AAGAAGTAGAAGAAACCCGTAGAGGGCGAAGAAACTGCAGAAGGAGAAAGATGGAAGAGG	2100
Qy	2101	tacctgcagaagtagaagaagtgggaagaggtacctgcagaagtagaagaagtgggaagag	2160
Db	2101	TACCTGCAGAAGTAGAAGAGTGAAGAGGTACCTGCAGAACTAGAGAGTGAAGAGG	2160

Qy	2161	taccagaagaagttagaagagggtaccgcgagaagtagaagaagttggaagagggtaccagaag	2220
Db	2161	TACCAGAAGAAGTAGAAGAGGTTACCCGCGAAGTAGTAGAANAAGTGGAGAAGGTACCAAG	2220
Qy	2221	aagtgaagagggtaccagaagaagtggaaagagggtaccagaagaagttggaagagggtaccag	2280
Db	2221	AGTGGAGAAGGTTACCAAGAAGAGTGGAAAGGTTACCAAGAGAAGTGGAGAAGGTACCAG	2280
Qy	2281	aagaagtggaaagaagtggaaagaagttagaagaagtagaggtaccagcggtagtagaagtag	2340
Db	2281	AAGAAGTGGAAAGAAGTGGAAAGAAGTAGAAGAAGTACAGGTTACCAAGCGGTAGTACAACTAG	2340
Qy	2341	aagtaaccagcggtagtagaagaagagggtccagaagaagtagaagaagaagaagaagagg	2400
Db	2341	AGGTACCACGCGTAGTAGAAGAAGAGGTTGCCAAGAANAAGTAGAAGAAGAACAAGAAGAGG	2400
Qy	2401	aagaaccagtagaggaagaagatgattacaattagtaataccatcggagaagaagatatc	2460
Db	2401	AGAACCCAGTAGAGGAAGAAGAAGTAGTATTACAAATTAGTAAATACCATCCGAGNAGAGTATAC	2460
Qy	2461	aattagacaaccagaagaagcgaattaggctctgggaattttatctatcatcgcacatgc	2520
Db	2461	AATTAGACAACCAAGAAAGACGAATTAGGCTCTGGAAATTTATCTATCATCTGCACATGC	2520
Qy	2521	actacaagaagcttccaaagggaatttatggaagaagaagaanaactgcagctgataccat	2580
Db	2521	ACTACCAAGACGTTCCAAAGGAATTTATGGAAAGAAGAAGAANAACATGCAGTGTATCCAT	2580
Qy	2581	tgaaccagaagaatttgcgaaggaagattcacaaattcacagaatgctcacattcattc	2640
Db	2581	TGAACCAAGAAGATTTGCAAGGAAGAAGTTCACAATCTACAGAATGGCTCAATTCAATTC	2640
Qy	2641	aaggcctagaagcgactgggaacgattagaagtagagcttaaaataaggctagagaagaat	2700
Db	2641	AAGGCCTAGAAGCGCAGTGGGAACGATTAGAAGTCAGCTTAAATAAGGCTAGAGAAAAGAT	2700
Qy	2701	ggatggaacaaagaataaagaataggctggcggctctgcttaattgaaaaataaaagggt	2760
Db	2701	GGATGGAAACAAGAATAAAGAAATGGGCTGGCTGGCTTCTGCTTAATTGAAAAATAAATGGT	2760
Qy	2761	cagaaatatagtcaaaatttcaacaaaggaaaggacccagctgggtttgagaaaaacgagagt	2820
Db	2761	CAGNATATAGTCAAAATTTCAACAAAAGNAAGGCCAGCCAGCTGGTTTGNAAAAACAGAGT	2820
Qy	2821	ggagcgacgagaagtggaaaaaatggtttaaagcagaagtcaaaatcccaaatgattcac	2880
Db	2821	GGAGCGACGAGAATCGAAAAAATGGTTTAAAGCAGAGAAGTCAAAATGCCAAAATTCATTCCAC	2880
Qy	2881	acttgaaaaaatgagtagcagcacatctccaatttatttaaaaattcttgtgaaagata	2940
Db	2881	ACTTGAAAAAANTGGATGAACGACACTCAATTCCAATTTATTTAAAAATTCCTTGAAAAAGATA	2940
Qy	2941	tgtcaacaattgaaacaagaacaccaaagaatggcttaatgaaactcactggaaaaagaacg	3000
Db	2941	TGTCACAAATTTGAAAAACAAGAAACCAAGAATGCTTAATGAATCACTCGGAAAAAAGAACG	3000
Qy	3001	aacgggggttatggctctgataatcatttgaagttatgaccacatcaaaattataaatgtgg	3060
Db	3001	AACGGGGTTATGGTCTCGAATCAATTTGAAGTTATGACCACATCAAAATTTATTAATCTGG	3060
Qy	3061	ctaagatcgagaatgtaccgtgcgaatctctaataataatagagaagaagagaactca	3120
Db	3061	CTAAGAGTCGAGAATGGTACCGTGGCAATTCCTTAATTAATAATAGAGAAGAAGAGAACTCA	3120
Qy	3121	tgaataggttctctctaaagaaaaacgaataattaggacaaagaatggaaaaaatggact	3180
Db	3121	TGAAATGGTTCTTCCTAAAAGAAAAACGAATATTTAGGACAAAGAATGCAAAAAATGGACT	3180
Qy	3181	cattggaaaaaagctaaatttttttggttcaattcaatgtgtacaaacatttctcgtgaaaa	3240
Db	3181	CATTGGAAAAAAGCTTAAATTTTTTGTGTCAATTCAAATGCTGTACAACTTTCTGTGGAAAA	3240

Qy	3241	cgcctaaccagaagaatgaatcaatttcttaagtataaaagtattgaattatgaaa	3300
Db	3241	CGCCTTAACCAAGGAAGAAATGGAATCAATTTGTTAATGAAATAAAAGTTTGAAATTATAGAA	3300
Qy	3301	aaagaacagattattctcttataaaataaataatc	3337
Db	3301	AAAAGAAGATTTCTCTTATAAAATAAATAATTC	3337
RESULT	2		
LOCUS	I23337	3337 bp	DNA linear PAT 07-OCT-1996
DEFINITION	Sequence 1 from patent US 5532133.		
ACCESSION	I23337		
VERSION	I23337.1	GI:1603207	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 3337)		
AUTHORS	Barnwell,J.W.		
TITLE	Plasmodium vivax blood stage antigens, PvESP-1, antibodies, and diagnostic assays		
JOURNAL	Patent: US 5532133-A 1 02-JUL-1996;		
FEATURES	Location/Qualifiers		
source	1..3337		
BASE COUNT	1304 a 467 c 875 g 691 t		
ORIGIN	/organism="unknown"		
Query Match	100.0%;	Score 3337;	DB 6; Length 3337;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 3337;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	gaattccgggtaaagtaacaacatggttctcgtagctatctatatataaacttactaatatttc	60
Db	1	GAATTCCGGTAAAGTAACAACATATGGTTTCGTATCTATATATAAACCCTTACTTAATTTTATC	60
Qy	61	tttgcgtttctttaaattcgtctcaacagtagaataaaaaataactataaaagctgc	120
Db	61	TTTTGCTTTTCTTTTAATTCATGCTTCANCAAGTAGATAAAAATAATCTATAAAACATGC	120
Qy	121	tatatatacatatatattcataagtgggcatttctgtaattgcgcatcatttaattacgta	180
Db	121	TATATATACATATATATTCTAAGTGGCATTTGTGAATTCGCATCATTTAAATTTACGTA	180
Qy	181	aaacaaatctgaaaaaaatt	240
Db	181	AAACAAATATTGAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTCTACAGAACGATTTAG	240
Qy	241	aattggaaaaagcttctctgatgatgttgtagagggtggagatcccttcaaacgacggtttag	300
Db	241	AATTGGAAAAATGCTTCTCATGATGTTGTAGAGGTGGAGGATCCTTCAAACGACGTTTAG	300
Qy	301	aattagaagagaaaaattttgatgagaatttcaggtgatgataaactcttttagatgcta	360
Db	301	AATTAGAAGAGGAAAAATTTTGATGAGAAATTCAGGTGATGATCAAACTCTTTTACATGCTA	360
Qy	361	cccccgaaatgactttgccttaacagatttgcceattgcaagcagatgaggaagtcaacy	420
Db	361	CCCCCGAAGATGACTTTTGCCTTAACAGATTTGCCAATTCAGACGATGAGGAAGTCAACG	420
Qy	421	aaacgtttagatggaggtgaaatcatataggagaggtttccactgaagatatggaacaagaag	480
Db	421	AAACGTTTAGATGGAGGTCAATCATTTAGCAGAGGTTTCCACCTGAAGATATGGAACAGAAAG	480
Qy	481	atggctcaacagatgatacgaacacagaaagagactacctgggtgatgataagagagaag	540
Db	481	ATGGCTCAACAGATGATACGGAACACGAAGAAGGACTACCTCGGTGATATGGAAGGAGAAAG	540
Qy	541	aagaagctggcgcatatggaagcagggggaagactggtgatttggaaagcgggggaagaaa	600

Db 541 AAGAGCTGGCGATATGGAAGCAGGGGAGAAAGCTGGTGAATTTGGAAGCAGGGGGAAGAAA 600
Qy 601 ctggcgatttgaagcaggggaagaaactggcgatttggaaagcaggggaagaaactggtg 660
Db 601 CTGGCGATTTGGAAGCAGGGGGAAGAACTGGCGATTTGGAAGCAGGGGGAAGAACTGGTG 660
Qy 661 atttgaagcaggggaagaaactggcgatttggaaagcaggggaagaaactggagatcg 720
Db 661 ATTtGGAAGCAGGGGGAAGAACTGGCGATTTGGAAGCAGGGGGAAGAACTGGAGATGGG 720
Qy 721 aaactgaagaagagcaactagagatcggaagaaactgaaatggaacaaactgtgtatgag 780
Db 721 AAACtGAAGAAGAGCAACTAGGAGATGCGGAACtGAAATtGGACAACTGTGTATGTAG 780
Qy 781 acacagaagatagttcagctgagcagaaagatcatgttctctctcaagaaatg 840
Db 781 ACACAGAAGATAGTTCAGCTGATGAGCAGAAAAAAGTACATGTTCTCTCAAGAAAAATG 840
Qy 841 tacaacctgcgatatagtaataatgacctcttggaaagtatttggataaaagataatt 900
Db 841 TACAACCTGCCGATAGTAATGATGCCCTCTTTGGAAGTATTTGGATAAAGATATAATTT 900
Qy 901 ttgataatataagatctcagccactattcgaaacaaatgtgctgctactgctaaac 960
Db 901 TTGATCATATTAAGATTTTCAGCCCACTATTTCACACAAATTTGGCGGGTACTGCTAAAC 960
Qy 961 atgttacgggacaagaattggcaatgaaacctgtaccattaccagtgccagagagcccg 1020
Db 961 ATGTTACGGGACAAGAATTTGCAATGAAACCTGTACCATTACCAGTGGCAGAAGGCCCG 1020
Qy 1021 ccaagtagtaccagcgaagaattagatgccactccagagagtagtactcgcattagatgta 1080
Db 1021 CCCAAGTACCAGCGGAAGAATTTAGATGCCACTCCAGAGAGTAGTTCGCATTAGATGTTA 1080
Qy 1081 cagaatctcccgaggaagtagaattagattagatgaagagcaactgaagaagaatcaa 1140
Db 1081 CAGAATCTCCGAGGAAGTAGAATTTAGTATTAGATGAGAGCAACTGAAGAGAATCAA 1140
Qy 1141 cggaaagtgggacaaacggaaggaagacaaacgaagaattagatgccactccagagatg 1200
Db 1141 CGGAAGTGGGACCAACGGGAAGAGACCAACCGAAGAAATTAGATGCCACTCCAGAGGATG 1200
Qy 1201 gatttcgcattagcgaactcgaagagaaacagagaacacagagaacacagaggaagaaa 1260
Db 1201 GATTTCGCATTAGACGAACCTCGAAGAGAGAAACAGAGAACCTAGAGGGGGAAGAAA 1260
Qy 1261 cagaagaactcgaagaggaagtatcagaagaactccagaagagaaagagattag 1320
Db 1261 CAGAAGAAGCTCCAGAGGAGAAGTATCAGAGAAGAACTCCAGAGGAGAGAAGAGTAG 1320
Qy 1321 aggcaactccagaggatgatttcgcattagatggaaactacattagaagaacccgaagaa 1380
Db 1321 AGGCCAACTCCAGAGCATGATTTCCGATTTAGATGGAACtACATTAGAAGAAACCGAAGAA 1380
Qy 1381 ctgcagaaggaagaacccgttagagggaagaacccgttagagggaagaagaacccgttag 1440
Db 1381 CTGCAGAAGGAGAAGAAACCGTAGAGGGAGAGAAGAAACCGTAGAGGGGAGAAGAAACCGTAG 1440
Qy 1441 agggagaagaactcgaagagagaagagatagaggaactccagagatgacttcc 1500
Db 1441 AGGGAGAGAAGCTCCAGAGAGAGAGAGAGATTTAGAGCAACTCCAGAGGATGACTTCC 1500
Qy 1501 aatagaagaacactcaggagaagagaaggggaaggaaggaaggaaggaaggaag 1560
Db 1501 AANTAGAGAACCATCAGGAGAAGGAGAAGGGGAAGGAGAGAGAGAGAGAGAGAGAG 1560
Qy 1561 gagaagcgttagtagcagtagtgcagtagtgcagtagtgcagtagtgcagtagtgcagtag 1620
Db 1561 GAGAAGCCGTAGTAGCAGTGCAGTAGTGCAGTAGTGCAGTAGTGCAGTAGTGCAGTGC 1620
Qy 1621 agcctgtcaaaccaatggtcgcctcaacgagagatgaaacttattcgttgatatttag 1680
Db 1621 AGCCTGTCAAACCAATGGTCTCCACGGCAGATGAACACTTATTTCGTTGATATCTTAG 1680

Qy 1681 ataagatttaacgtatgcagacattacatcctttgagccattattttaacaaactcctca 1740
Db 1681 ATAACGATTTAACGATATGCAGACATTTACATCCTTTTGAGCCATTTATTTAAACAAATCCTCA 1740
Qy 1741 aggatccttgatgcagagagagcgttaacagttaccattcaaaaggaagcactgtacaagtac 1800
Db 1741 AGGATCCTTGATGCGAGGAGAGGCTGTAAACAGTACCATCAAGGAAGACCTGTACAAAGTAC 1800
Qy 1801 cagtggcagtagggcccgcaagagtgccaacggaagaaattgatgcaactcccaagagg 1860
Db 1801 CAGTGGCAGTAGGGCCCGCAAGAGTGGCCACCGAAGAAATTGATGCAACTCCCAAGAGG 1860
Qy 1861 acgatttcgaattagaaggaactgcagaagctccagaggaaggaagaaattagattagaag 1920
Db 1861 ACGATTTCCAATTTAGAGGAACtGCAGAAGCTCCAGAGGAAGGAGAATTTAGTATTAGAAG 1920
Qy 1921 gagaaggaagaaaccaaaggaagcaagagaaggaagcaacagaaggaagagatgc 1980
Db 1921 GAGAAGGAGAACCACCGGAAGAAAGCAAGAGAGAGAGCCACACAGGAAGGAAGTGC 1980
Qy 1981 cagaagaagaattagagcgaactccagagagcgaatttcgaaattagaagaaccaaagag 2040
Db 1981 CAGAAGAAGATTTAGAGGCAACTCCAGAGGAGATTTCSAATTTAGAAGAACCAACAGGAG 2040
Qy 2041 aagaagtagaagaacccgttagaggcggaagaaactcgaaggaaggaagaaagtggaaaggg 2100
Db 2041 AAGAAGTAGAAGAAACCTTAGAGGCGGAAGAACTCCAGAGGAGAGAAGTGGGAAGAGG 2100
Qy 2101 tacttcgagaagtagaagaagtggaaaggttacctcgagaagtagaagaagtggaaaggg 2160
Db 2101 TACCTGCAAGAGTAGAAGAAGTGGAAAGGTACCTTCGAGAGTAGAAGAAGTGGGAAGAGG 2160
Qy 2161 taccagaagaagtagaagaggttaccgcgagaagtagaagaagtggaaaggttaccagaag 2220
Db 2161 TACCAGAAGAAGTAGAAGAGGTACCCGAGAGAGTAGAAGAAGTGGGAAGAGGTACCAGAAG 2220
Qy 2221 aagtggaaaggttaccagaagaagtggaaaggttaccgaaggaaggtggaaagaggttaccag 2280
Db 2221 AAGTGGAAAGAGGTACCAGAAGAAGTGGAAAGGTACCAGAGAGAGTGGGAAGAGGTACCAG 2280
Qy 2281 aagaagtggaaagaagtggaaagaagtagaagaagtagaggttaccgcggttagtagaagtag 2340
Db 2281 AAGAAGTGGAGAAGTGGAGAAGTAGAAGAAGTAGAAGTACCAGCGGTAGTAGAAGTAG 2340
Qy 2341 aagtagccgcggttagtagaagaagaggtgccagaagaagtagaagaagaagaagaagagg 2400
Db 2341 AAGTACCAGCGGTAGTAGAAGAAGAGGTGCCAGAAGAAGTAGAAGAAGGAAGAAGAGG 2400
Qy 2401 aagaaccagtagagaagaagaatgtattacaattagtaataaccatcggaagaagataatc 2460
Db 2401 AAGAACCAGTAGAGAAGAAGATGTATTACAATTTAGTATTACCATCGGAAGAAGATATAC 2460
Qy 2461 aattagcaaaccaagaagaacgaattaggtctctggaaatttctatctatctgcagatgc 2520
Db 2461 AATTAGACAACCAAGAAGAACGAATTAGGCTCTGGAAATTTTATCTATCATCGACATGC 2520
Qy 2521 actaccaagaagttcccaaggaatttatggaaagaagaagaactgcagtgatccat 2580
Db 2521 ACTACCAAGACGTTCCAAAAGGAATTTATGGAAAGAAGGAAGAAACTGCAGTGTATCCAT 2580
Qy 2581 tgaaccagaagatttgcaaaaggaagattcaaatctacagaaatggctcaactcattc 2640
Db 2581 TGAACCAAGAGAATTTTCAAGAGGAAGATTTCAACATCTACAGAATTTGGCTCAGATTTCTTC 2640
Qy 2641 aagccttagaagggcagctggaaacattagaagtgcgttaaatagagcttagaagaagat 2700
Db 2641 AAGGCTTAGAAGGCCAGCTGGGAACGATTTAGAAGTGAAGCTTAATTAAGGCTTAGAGAAGAT 2700
Qy 2701 ggaatgaagaagaataaagaatgggctgcgtcttcgcttaattgaaataaattggt 2760
Db 2701 GGATGGAACAAAGAAATAAAGAAATGGGCTGGCTTCGCTTAATTTGAAAAATAAATGGT 2760

```
Qy 2761 cagaatactagtcacaaatttcaacaaaggaagagccagctggttttgagaaacagagagt 2820
Db 2761 CAGATATAGTCAATTTTCAACAAAGGAAGAGCCAGCTGGTTTGAGAAACGAGAGT 2820
Qy 2821 ggaagcagcagaataatggaacaaatggtttaagcagaagtcacaaatcccaaatgattcac 2880
Db 2821 GCAGCGAGGAGAAATGGAAGAAATGCTTTAAAGCAGAGTCAAAATCCCAAAATGATTCCAC 2880
Qy 2881 acttgaacaaatgatgaacacactcattcccaatttatttataaattcttgtgaagata 2940
Db 2881 ACTTGAAGAAATGATGAACACACTCAATCCAAATTTATTAAATTTCTGTGAAGATA 2940
Qy 2941 tgtcacaaattgaaacaaagaaaccaaagaaatggttaataatcactggaacaaagaacg 3000
Db 2941 TGTCACAATTTGAAACAAAGAAACCAAGAAATGTTAATCAATCACTGGGAAAGAACG 3000
Qy 3001 aacggggtatgggttctgaacatttggaagttagaccacatcaaaattatttaattgtgg 3060
Db 3001 AACGGGGTATGGGTCTGAATCATTTTGAAGTTATGACCACATCAAAATTTATTAATGTGG 3060
Qy 3061 ctaagagtcgagaatggtaccgtgccatcctaataataatagagaagaagagaactca 3120
Db 3061 CTACAGCTCGAGATGTTACCGTCCCAATCTATATATAATAGAGAAAGAGAACTCA 3120
Qy 3121 tgaatggttctcctcaaaagaaacgaataatttaggacaaagaatggaaaaatggagct 3180
Db 3121 TGAATGTTTTCTCTAAACAAAGCAATATTTAGGACAAAGAAATGGAATAATGGACT 3180
Qy 3181 catggaaaaaagttaatttttgggttcacattcaatcgaatggtacacaaatcttctggaaaa 3240
Db 3181 CATTTGAAAAAAGTTAAATTTTGTGTTCAATTTGTTACAACTGTTACAACTTTTCTGGAAA 3240
Qy 3241 cgcctaaccaaggagaatgaatacttctgaataataataaagttgaattatgaata 3300
Db 3241 CGCCTAACCAAGGAGAGATGAATCAATTTGTTAATGAATAAAGTTTGAATTATAGAA 3300
Qy 3301 aaagaacagattattctcttataaaaaataaattc 3337
Db 3301 AAAAGAACAGATTATCTCTTATAAAATAAATAATTC 3337

RESULT 3
AC067910/c
LOCUS
DEFINITION Homo sapiens chromosome 22 clone RP11-51H9 map 22, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC067910
VERSION AC067910.1 GI:7652049
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 70206)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 22, clone RP11-51H9
Unpublished
2 (bases 1 to 70206)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campolano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hegos,B., Hearford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieue,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
```

TITLE
JOURNAL

COMMENT

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,W., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7007
Center clone name: 51_H_9

* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 736: contig of 736 bp in length
* 737 836: gap of 100 bp
* 837 1561: contig of 725 bp in length
* 1562 1661: gap of 100 bp
* 1662 2378: contig of 717 bp in length
* 2379 2478: gap of 100 bp
* 2479 3202: contig of 724 bp in length
* 3203 3302: gap of 100 bp
* 3303 4006: contig of 704 bp in length
* 4007 4106: gap of 100 bp
* 4107 4798: contig of 692 bp in length
* 4799 4898: gap of 100 bp
* 4899 5611: contig of 713 bp in length
* 5612 5711: gap of 100 bp
* 5712 6416: contig of 705 bp in length
* 6417 6516: gap of 100 bp
* 6517 7213: contig of 697 bp in length
* 7214 7313: gap of 100 bp
* 7314 8017: contig of 704 bp in length
* 8018 8117: gap of 100 bp
* 8118 8833: contig of 716 bp in length
* 8834 8933: gap of 100 bp
* 8934 9656: contig of 723 bp in length
* 9657 9756: gap of 100 bp
* 9757 10472: contig of 716 bp in length
* 10473 10572: gap of 100 bp
* 10573 11284: contig of 712 bp in length
* 11285 11384: gap of 100 bp
* 11385 12108: contig of 724 bp in length
* 12109 12208: gap of 100 bp
* 12209 12910: contig of 702 bp in length
* 12911 13010: gap of 100 bp
* 13011 13710: contig of 700 bp in length
* 13711 13810: gap of 100 bp
* 13811 14500: contig of 690 bp in length
* 14501 14600: gap of 100 bp
* 14601 15305: contig of 705 bp in length
* 15306 15405: gap of 100 bp
* 15406 16114: contig of 709 bp in length
* 16115 16214: gap of 100 bp
* 16215 16914: contig of 700 bp in length

```
* 16915 17014: gap of 100 bp
* 17015 17722: contig of 708 bp in length
* 17723 17822: gap of 100 bp
* 17823 18517: contig of 695 bp in length
* 18518 18617: gap of 100 bp
* 18618 19329: contig of 712 bp in length
* 19330 19429: gap of 100 bp
* 19430 20132: contig of 703 bp in length
* 20133 20232: gap of 100 bp
* 20233 20955: contig of 723 bp in length
* 20956 21055: gap of 100 bp
* 21056 21762: contig of 707 bp in length
* 21763 21862: gap of 100 bp
* 21863 22560: contig of 698 bp in length
* 22561 22660: gap of 100 bp
* 22661 23351: contig of 691 bp in length
* 23352 23451: gap of 100 bp
* 23452 24145: contig of 694 bp in length
* 24146 24245: gap of 100 bp
* 24246 24952: contig of 707 bp in length
* 24953 25052: gap of 100 bp
* 25053 25756: contig of 704 bp in length
* 25757 25856: gap of 100 bp
* 25857 26568: contig of 712 bp in length
* 26569 26668: gap of 100 bp
* 26669 27391: contig of 723 bp in length
* 27392 27491: gap of 100 bp
* 27492 28187: contig of 696 bp in length
* 28188 28287: gap of 100 bp
* 28288 28993: contig of 706 bp in length
* 28994 29093: gap of 100 bp
* 29094 29799: contig of 706 bp in length
* 29800 29899: gap of 100 bp
* 29900 30629: contig of 730 bp in length
* 30630 30729: gap of 100 bp
* 30730 31418: contig of 689 bp in length
* 31419 31518: gap of 100 bp
* 31519 32200: contig of 682 bp in length
* 32201 32300: gap of 100 bp
* 32301 33002: contig of 702 bp in length
* 33003 33102: gap of 100 bp
* 33103 33782: contig of 680 bp in length
* 33783 33882: gap of 100 bp
* 33883 34581: contig of 699 bp in length
* 34582 34681: gap of 100 bp
* 34682 35382: contig of 701 bp in length
* 35383 35482: gap of 100 bp
* 35483 36178: contig of 696 bp in length
* 36179 36278: gap of 100 bp
* 36279 36998: contig of 720 bp in length
* 36999 37098: gap of 100 bp
* 37099 37805: contig of 707 bp in length
* 37806 37905: gap of 100 bp
* 37906 38620: contig of 715 bp in length
* 38621 38720: gap of 100 bp
* 38721 39424: contig of 704 bp in length
* 39425 39524: gap of 100 bp
* 39525 40243: contig of 719 bp in length
* 40244 40343: gap of 100 bp
* 40344 41048: contig of 705 bp in length
* 41049 41148: gap of 100 bp
* 41149 41851: contig of 703 bp in length
* 41852 41951: gap of 100 bp
* 41952 42658: contig of 707 bp in length
* 42659 42758: gap of 100 bp
* 42759 43478: contig of 720 bp in length
* 43479 43578: gap of 100 bp
* 43579 44282: contig of 704 bp in length
* 44283 44382: gap of 100 bp
* 44383 45080: contig of 698 bp in length
* 45081 45180: gap of 100 bp
* 45181 45894: contig of 714 bp in length
* 45895 45994: gap of 100 bp
```

```
* 45995 46707: contig of 713 bp in length
* 46708 46807: gap of 100 bp
* 46808 47546: contig of 739 bp in length
* 47547 47646: gap of 100 bp
* 47647 48350: contig of 704 bp in length
* 48351 48450: gap of 100 bp
* 48451 49166: contig of 716 bp in length
* 49167 49266: gap of 100 bp
* 49267 49978: contig of 712 bp in length
* 49979 50078: gap of 100 bp
* 50079 50785: contig of 707 bp in length
* 50786 50885: gap of 100 bp
* 50886 51582: contig of 697 bp in length
* 51583 51682: gap of 100 bp
* 51683 52384: contig of 702 bp in length
* 52385 52484: gap of 100 bp
* 52485 53196: contig of 712 bp in length
* 53197 53296: gap of 100 bp
* 53297 54015: contig of 719 bp in length
* 54016 54115: gap of 100 bp
* 54116 54833: contig of 718 bp in length
* 54834 54933: gap of 100 bp
* 54934 55620: contig of 687 bp in length
* 55621 55720: gap of 100 bp
```

Query Match 1.2%; Score 40; DB 2; Length 70206;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1527 gaaggggaagagagagagagagagagagagagagagag 1566
|||||
Db 24332 GAAGGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 24293

RESULT 4

AC027625 AC027625 165868 bp DNA linear HTG-09-AUG-2001
LOCUS Homo sapiens chromosome 4 clone RP11-27013, WORKING DRAFT SEQUENCE,
DEFINITION 2 unordered pieces.
ACCESSION AC027625.4 GI:14333966
VERSION AC027625.4
KEYWORDS HTG; HTGS_P1ASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 165868)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165868)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 9, 2001 this sequence version replaced gi:9838122.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0270103
----- Summary Statistics -----
Sequencing vector: M13, 40%
Chemistry: Dye-primer ET; 40% of reads
Chemistry: Dye-terminator Big Dye; 60% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165641 bases at least Q40
Consensus quality: 165729 bases at least Q30

Consensus quality: 165743 bases at least Q20
 Insert size: 168000; agarose-fp
 Sequencing vector: 166538; sum-of-contigs
 Quality coverage: 11.53 in Q20 bases; agarose-fp
 Quality coverage: 11.63 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 57897: contig of 57897 bp in length
 * -57898 57997: gap of unknown length
 * 57998 165868: contig of 107871 bp in length.

FEATURES

source
 Location/Qualifiers
 1..165868
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /clone="RP11-27013"

misc_feature

1..57897
 /note="assembly_name:Contig5
 clone_end:SP6
 vector_side:right"

misc_feature

57998..165868
 /note="assembly_name:Contig6
 clone_end:T7
 vector_side:left"

BASE COUNT 48495 a 32516 c 32464 g 52293 t 100 others
 ORIGIN

Query Match 1.2%; Score 40; DB 2; Length 165868;
 Best Local Similarity 100.0%; Pred. No. 2.9e-08;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1527 gaagggagagagagagagagagagagagagagag 1566
 |||||
 Db 3740 GAAGGGGAGGAGAGGAGAGGAGGAGGAGGAGAG 3701

RESULT 5
 AL451079 186562 bp DNA linear HTG 05-OCT-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-307017, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 2 unordered pieces.

ACCESSION AL451079 AC025875
 VERSION AL451079.6 GI:15983906
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 186562)
 WILLIAMS.S.

Direct Submission
 Submitted (04-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerquest@sanger.ac.uk
 On Oct 8, 2001 this sequence version replaced gi:15591231.
 Draft Sequence Produced by Whitehead Institute/MIT Center for
 Genome Research, 320 Charles Street,
 Cambridge, MA 02141, USA
 http://www-seq.wi.mit.edu

Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information

Center project name: BA307017
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: M13; M7815; 26% of reads
 Chemistry: Dye-terminator Big Dye; 108752; 73% of reads
 Consensus quality: 185620 bases at least Q40
 Consensus quality: 185985 bases at least Q30
 Consensus quality: 186189 bases at least Q20
 Insert size: 18462; sum-of-contigs
 Insert size: 172836; 9.6% error; agarose-fp
 Quality coverage: 12.37x in Q20 bases; sum-of-contigs Quality
 coverage: 13.42x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 2201: contig of 2201 bp in length
 * 2202 2301: gap of 100 bp
 * 2302 186562: contig of 184261 bp in length.

FEATURES

source
 Location/Qualifiers
 1..186562

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-307017"
 /clone_lib="RPC1-11.2"
 1..2201

misc_feature

/note="assembly_fragment:00529"
 2302..186562
 /note="assembly_fragment:05314
 clone_end:T7
 vector_side:right"

BASE COUNT 50258 a 43749 c 43919 g 48534 t 102 others
 ORIGIN

Query Match 1.0%; Score 34; DB 2; Length 186562;
 Best Local Similarity 100.0%; Pred. No. 2e-05;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1533 gaagagagagagagagagagagagagagagag 1566
 |||||
 Db 27017 GAAGGAGAGGAGAGGAGGAGGAGGAGGAGAGAG 27050

RESULT

AC096541/c 6
 LOCUS AC096541 200898 bp DNA linear HTG 18-SEP-2001
 DEFINITION Homo sapiens chromosome 1 clone RP11-330M19, WORKING DRAFT
 SEQUENCE, 4 unordered pieces.

ACCESSION AC096541 AL359084

VERSION AC096541.1 GI:15638689

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 200898)

AUTHORS Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J. and Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 200898)
 AUTHORS Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J. and Haugen,E.D.
 TITLE Direct Submission

Submitted (18-SEP-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Sep 18, 2001 this sequence version replaced gi:12331065.
 COMMENT

----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgctgs@u.washington.edu
 Drafting Center: SC
 ----- Project Information
 Center project name: chr-1
 Center clone name: RP11-330M19 (sc0667)
 ----- Summary Statistics

Sequencing vector: plasmid; 38% of reads
 Sequencing vector: plasmid; 108752; 62% of reads
 Chemistry: Dye-terminator ET; 89% of reads
 Chemistry: Dye-terminator Big Dye; 11% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 199175 bases at least Q40
 Consensus quality: 199939 bases at least Q30
 Consensus quality: 200341 bases at least Q20
 Insert size: 200598; sum-of-contigs
 Quality coverage: 9.5x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 2137: contig of 2137 bp in length
 * 2138 2237: gap of unknown length
 * 2238 29137: contig of 26900 bp in length
 * 29138 29237: gap of unknown length
 * 29238 55080: contig of 25843 bp in length
 * 55081 55180: gap of unknown length
 * 55181 200898: contig of 145718 bp in length.

FEATURES

source
 1. .200898
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-330M19"
 /clone_lib="RPCI human BAC library 11"
 misc_feature
 1. .2137
 /note="assembly_name:Contig18"
 misc_feature
 2238. .29137
 /note="assembly_name:Contig19"
 misc_feature
 29238. .55080
 /note="assembly_name:Contig20"
 misc_feature
 55181. .200898
 /note="assembly_name:Contig21"
 BASE COUNT 50081 a 49279 c 48900 g 52334 t 304 others
 ORIGIN

Query Match 1.0%; Score 34; DB 2; Length 200898;
 Best Local Similarity 100.0%; Pred. No. 2e-05;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1533 gaaggagaaggaggggaaggagaaggagaag 1566
 |||
 Db 55487 GAAGGAGAGGAGAGGGGAGGAGGAGGAGAG 55454

RESULT

AL645546/c 7
 LOCUS AL645546 225609 bp DNA linear HTG 13-NOV-2001
 DEFINITION Mus musculus chromosome 13 clone RP23-124G18, *** SEQUENCING IN
 PROGRESS ***, in unordered pieces.
 ACCESSION AL645546
 VERSION AL645546.4 GI:16944328
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (sites)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
 Submitted (08-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Nov 15, 2001 this sequence version replaced gi:16754763.

----- Genome Center

Center: UK Medical Research Council

Center code: UK-MRC

Web site: <http://mrcseq.har.mrc.ac.uk>

Contact: mouseq@har.mrc.ac.uk

----- Project Information

Center project name: BM124G18

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; 108752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 224286 bases at least Q40

Consensus quality: 224563 bases at least Q30

Consensus quality: 224933 bases at least Q20

Insert size: 225409; sum-of-contigs

Insert size: 216124; 7.8% error; agarose-fp

Quality coverage: 11.91x in Q20 bases; sum-of-contigs Quality

coverage: 13.08x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

FEATURES

source

1. .225609

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="13"

/clone="RP23-124G18"

/clone_lib="RPCI-23"

1. .101657

/note="assembly_fragment:01246"

clone_end:T7

vector_side:left

101758. .190829

/note="assembly_fragment:02144.0"

190930. .225609

/note="assembly_fragment:03429"

clone_end:SP6

vector_side:right

BASE COUNT 66198 a 45409 c 44739 g 69047 t 216 others

ORIGIN

Query Match 1.0%; Score 34; DB 2; Length 225609;

Best Local Similarity 100.0%; Pred. No. 1.9e-05;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1533 gaaggagaaggaggggaaggagaaggagaag 1566

|||||

Db 101565 GAAGGAGAAGGAGAGGGGAGGAGGAGGAGAG 101532

RESULT

AX251053

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AX251053 6862 bp DNA linear PAT 05-OCT-2001

Sequence 21 from Patent WO0168912.

AX251053 GI:15984476

synthetic construct.

synthetic construct

artificial sequence.

REFERENCE	1 (bases 1 to 6862)
AUTHORS	Olek.A., Piepenbrock,C. and Berlin,K.
TITLE	Diagnosis of diseases associated with tumor suppressor genes and oncogenes
JOURNAL	Patent: WO 0168912-A 21 20-SEP-2001;
FEATURES	Epigenomics AG (DE)
source	Location/Qualifiers
	1..6862
	/organism="synthetic construct"
	/db_xref="taxon:32630"
	/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT	1278 a 518 c 2176 g 2890 t
ORIGIN	
Query Match	1.0%; Score 33; DB 6; Length 6862;
Best Local Similarity	100.0%; Pred. No. 9e-05;
Matches	33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	192 gaaaaaaattttttttttttttttttttttttttttttgtt 224
Db	853 GAAAAAAATTTTTTTTTTTTTTTTTTTTGTGTT 885
RESULT 9	
LOCUS	AX251776 6862 bp DNA linear PAT 05-OCT-2001
DEFINITION	Sequence 37 from Patent WO0168911.
ACCESSION	AX251776
VERSION	AX251776.1 GI:15985131
KEYWORDS	.
SOURCE	synthetic construct. artificial sequence.
ORGANISM	artificial sequence.
REFERENCE	1 (bases 1 to 6862)
AUTHORS	Olek.A., Piepenbrock,C. and Berlin,K.
TITLE	Diagnosis of diseases associated with the cell cycle
JOURNAL	Patent: WO 0168911-A 37 20-SEP-2001;
FEATURES	Epigenomics AG (DE)
source	Location/Qualifiers
	1..6862
	/organism="synthetic construct"
	/db_xref="taxon:32630"
	/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT	1278 a 518 c 2176 g 2890 t
ORIGIN	
Query Match	1.0%; Score 33; DB 6; Length 6862;
Best Local Similarity	100.0%; Pred. No. 9e-05;
Matches	33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	192 gaaaaaaattttttttttttttttttttttttttttttgtt 224
Db	853 GAAAAAAATTTTTTTTTTTTTTTTTTTTGTGTT 885
RESULT 10	
LOCUS	AX345124 6862 bp DNA linear PAT 01-FEB-2002
DEFINITION	Sequence 195 from Patent WO0200928.
ACCESSION	AX345124
VERSION	AX345124.1 GI:18493010
KEYWORDS	.
SOURCE	synthetic construct. artificial sequence.
ORGANISM	artificial sequence.
REFERENCE	1 (sites)
AUTHORS	Olek.A., Piepenbrock,C. and Berlin,K.
TITLE	Diagnosis of diseases associated with the immune system
JOURNAL	Patent: WO 0200928-A 195 03-JAN-2002;
FEATURES	Epigenomics AG (DE)
	Location/Qualifiers

source	1..6862
	/organism="synthetic construct"
	/db_xref="taxon:32630"
	/note="Chemically treated genomic DNA (Homo sapiens)"
BASE COUNT	1278 a 518 c 2176 g 2890 t
ORIGIN	
Query Match	1.0%; Score 33; DB 6; Length 6862;
Best Local Similarity	100.0%; Pred. No. 9e-05;
Matches	33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	192 gaaaaaaattttttttttttttttttttttttttttttgtt 224
Db	853 GAAAAAAATTTTTTTTTTTTTTTTTTTTGTGTT 885
RESULT 11	
HSA271735	HSA271735 240000 bp DNA linear PRI 07-JUL-2000
LOCUS	Homo sapiens Xq pseudoautosomal region; segment 1/2.
DEFINITION	AJ271735 Y19189
ACCESSION	AJ271735.1 GI:8979788
VERSION	
KEYWORDS	repeat region; repetitive DNA; sprouty (Drosophila) homolog 3; SPRY3 gene.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 240000)
	Ciccodicola,A., D'Esposito,M., Esposito,T., Gianfrancesco,F., Migliaccio,C., Milano,M.G., Matarazzo,M.R., Vacca,M., Franze,A., Cuccurese,M., Cocchia,M., Curci,A., Terracciano,A., Torino,A., Cocchia,S., Mercadante,G., Pannone,E., Archidiacono,N., Rocchi,M., Schlessinger,D. and D'Urso,M.
TITLE	Differentially regulated and evolved genes in the fully sequenced Xq/q pseudoautosomal region
JOURNAL	Hum. Mol. Genet. 9 (3), 395-401 (2000)
MEDLINE	2012249
REFERENCE	2 (bases 1 to 240000)
AUTHORS	Ciccodicola,A.
TITLE	Direct Submission
JOURNAL	Submitted (28-JUL-1999) A. Ciccodicola, International Institute of Genetics and, Biophysics (IIGB-CNR), Via P. Castellino 111, Napoli, 80131, ITALY
FEATURES	Location/Qualifiers
source	1..240000
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="Xq28"
repeat_region	1..176
	/note="DXS1108"
repeat_region	33..83
	/note="(AC) d"
repeat_region	96..157
	/note="HERVL d"
repeat_region	160..678
	/note="MLTH c"
repeat_region	768..1074
	/note="HERVL d"
repeat_region	1084..1214
	/note="Aluy d"
repeat_region	1232..1515
	/note="HERVL d"
repeat_region	1546..1716
	/note="HERVL d"
repeat_region	1718..1912
	/note="Alusc d"
repeat_region	1983..2379
	/note="HERVL d"
repeat_region	2433..2662
	/note="MER41I c"

repeat_region	2667. .2733	/note="MER21B d"
repeat_region	2737. .3281	/note="(AT) d"
repeat_region	3303. .3658	/note="L1PA7 c"
repeat_region	3663. .3760	/note="L1PA2 d"
repeat_region	4290. .4490	/note="HERVL d"
repeat_region	4701. .5003	/note="X20"
repeat_region	5120. .5260	/note="L1PA2 d"
repeat_region	5267. .5702	/note="L1R50 d"
repeat_region	5799. .5939	/note="L1R50 d"
repeat_region	5990. .6123	/note="L1MD2 c"
repeat_region	6515. .6953	/note="AluJo d"
repeat_region	6962. .7075	/note="L1R33 d"
repeat_region	7123. .7183	/note="L1MB7 c"
repeat_region	7200. .7402	/note="L1MA10 c"
repeat_region	7403. .7868	/note="L1MB7 c"
repeat_region	7869. .8160	/note="L1MA9 c"
repeat_region	8165. .8248	/note="MER28 c"
repeat_region	8249. .9160	/note="TIGGER2 c"
repeat_region	9161. .9833	/note="L1P_MA2 c"
repeat_region	9857. .9900	/note="L1R30 c"
repeat_region	9939. .11947	/note="L1P_MA2 c"
repeat_region	11949. .12598	/note="L1P_MA2 c"
repeat_region	12606. .13027	/note="L1P_MA2 c"
repeat_region	13032. .13675	/note="L1P_MA2 c"
repeat_region	13701. .14702	/note="L1P_MA2 c"
repeat_region	14859. .15042	/note="L1P_MA2 c"
repeat_region	15082. .15964	/note="L1P_MA2 c"
repeat_region	15965. .16512	/note="L1PA7 c"
repeat_region	16513. .18088	/note="L1 c"
repeat_region	18637. .18842	/note="L1 d"
repeat_region	18851. .19123	/note="L1MD2 c"
repeat_region	19127. .19334	/note="AluJb c"
repeat_region	19472. .19602	/note="L1MD2 c"
repeat_region	19610. .19995	/note="L1MA2 c"
repeat_region	19996. .20602	/note="MSTA c"
repeat_region	20607. .20882	/note="L1MA2 c"
repeat_region	21614. .21910	/note="L1MB7 c"
repeat_region	21925. .22090	/note="L1MA2 c"
repeat_region	22397. .22528	/note="X25"
repeat_region	23028. .23108	/note="L1P_MA2 c"
repeat_region	23109. .23479	/note="L1 c"
repeat_region	23481. .24297	/note="L1 c"
repeat_region	24299. .25822	/note="L1P5A1 c"
repeat_region	25829. .26690	/note="L1P_MA2 c"
repeat_region	26915. .27203	/note="AluY c"
repeat_region	27206. .27535	/note="L1 c"
repeat_region	27538. .28856	/note="L1P_MA2 c"
repeat_region	28860. .29705	/note="L1M2_5 c"
repeat_region	29716. .29923	/note="L1PBA_5 c"
repeat_region	29934. .30218	/note="L1PBB_5 c"
repeat_region	30226. .30307	/note="L1PA16_5 c"
repeat_region	30442. .30465	/note="(A) d"
repeat_region	31382. .31556	/note="MER5A d"
repeat_region	31726. .31912	/note="L1MA2 c"
repeat_region	32492	/note="32667. .X08"
repeat_region	32933. .33064	/note="L1ME3A d"
repeat_region	33072. .33390	/note="L1ME4 d"
repeat_region	33770. .34007	/note="LINE2 c"
repeat_region	34042. .34078	/note="L1MA5 c"
repeat_region	34083. .34374	/note="LINE2 c"
repeat_region	34467. .35368	/note="LINE2 c"
repeat_region	35491. .36126	/note="LINE2 c"
repeat_region	36127. .36352	/note="L1ME3A d"
repeat_region	37739. .37981	/note="L1ME2 d"
repeat_region	38272. .38364	/note="MIR c"
repeat_region	39305. .39363	/note="MIR d"
repeat_region	39436. .39991	/note="MER511 d"
repeat_region	39992. .40087	/note="L1 c"
repeat_region	40088. .40978	/note="L1 d"
repeat_region	42134. .42302	/note="L1PA2 d"
repeat_region	44358. .44681	/note="L3 c"
repeat_region	46558. .46702	/note="MSTA1 c"
repeat_region	46720. .46960	/note="LINE2 c"
repeat_region		/note="L1PA7 d"

Query Match 1.0%; Score 33; DB 9; Length 240000;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1527 gaagggaaggaaggaaggaaggaaggaagaa 1559

Db 190008 GAAGGGAAGGAAGGAAGGAAGGAAGGAAG 190040

RESULT 12

AC096301

LOCUS

DEFINITION Rattus norvegicus chromosome Rf1 clone CH230-233D20, WORKING DRAFT

AC096301 SEQUENCE, 33 unordered pieces.

AC096301 AC096301.4 GI:17943998

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 246176)

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Denn-Arroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J.J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguy,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,F., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Thomas,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williams,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 246176)

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:17062231.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GEWU

Center clone name: CH230-233D20

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 227668 bases at least Q40

Consensus quality: 231229 bases at least Q30

Consensus quality: 233891 bases at least Q20

Estimated insert size: 229594; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 4.9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 33 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 25258: contig of 25258 bp in length

* 25259 25358: gap of unknown length

* 25359 48246: contig of 22888 bp in length

* 48247 48346: gap of unknown length

* 48347 66685: contig of 18339 bp in length

* 66686 66785: gap of unknown length

* 66786 83731: contig of 16946 bp in length

* 83732 83831: gap of unknown length

* 83832 93990: contig of 15559 bp in length

* 93991 94990: gap of unknown length

* 94991 110000: contig of 10510 bp in length

* 110001 122163: contig of 12063 bp in length

* 122164 122263: gap of unknown length

* 122264 134794: contig of 12531 bp in length

* 134795 134894: gap of unknown length

* 134895 146779: contig of 11884 bp in length

* 146780 146878: gap of unknown length

* 146879 155054: contig of 8176 bp in length

* 155055 155154: gap of unknown length

* 155155 162599: contig of 7445 bp in length

* 162600 172264: contig of 9565 bp in length

* 172265 172364: gap of unknown length

* 172365 180466: contig of 8102 bp in length

* 180467 180567: gap of unknown length

* 180568 186883: contig of 6317 bp in length

* 186884 186983: gap of unknown length

* 186984 193290: contig of 6307 bp in length

* 193291 193390: gap of unknown length

* 193391 200294: contig of 6904 bp in length

* 200295 200394: gap of unknown length

* 200395 205791: contig of 5397 bp in length

* 205792 205891: gap of unknown length

* 205892 210659: contig of 4767 bp in length

* 210660 210759: gap of unknown length

* 210760 215486: contig of 4728 bp in length

* 215487 215586: gap of unknown length

* 215587 220044: contig of 4458 bp in length

* 220045 220144: gap of unknown length

* 220145 223051: contig of 2907 bp in length

* 223052 223152: gap of unknown length

* 223153 225621: contig of 2469 bp in length

* 225622 228107: gap of unknown length

* 228108 228207: gap of unknown length

```
* 228208 231207: contig of 3000 bp in length
* 231208 231307: gap of unknown length
* 231308 232936: contig of 1629 bp in length
* 232937 233036: gap of unknown length
* 233037 234444: contig of 1408 bp in length
* 234445 234544: gap of unknown length
* 234545 236460: contig of 1916 bp in length
* 236461 236560: gap of unknown length
* 236561 238859: contig of 2299 bp in length
* 238860 238959: gap of unknown length
* 238960 240759: contig of 1800 bp in length
* 240760 240859: gap of unknown length
* 240860 242544: contig of 1685 bp in length
* 242545 242644: gap of unknown length
* 242645 243695: contig of 1051 bp in length
* 243696 243795: gap of unknown length
* 243796 244851: contig of 1056 bp in length
* 244852 244951: gap of unknown length
* 244952 246176: contig of 1225 bp in length.
FEATURES
    source
    1..246176
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /chromosome="9L"
    /clone="CH230-233D20"
BASE COUNT 65649 a 53496 c 55272 g 66413 t 3346 others
ORIGIN

Query Match      1.0%  Score 33;  DB 2;  Length 246176;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1534 aaggagaaggaagggaagggaaggagaag 1566
|||||
Db 191928 AAGGAGAGGAGAGGAGGAGGAGGAGAG 191960

RESULT 13
AP001848
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-814018 map 18p11.2, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION AP001848
VERSION AP001848.3 GI:9757461
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-814018.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 224,010 genomic DNA of 18p11.2
Published Only in Database (2000) In press
2 (bases 1 to 224010)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (18-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Aug 9, 2000 this sequence version replaced gi:8117511.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-814018

----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads Assembly
Program: Phrap; version 0.990329
Consensus quality: 212358 bases at least Q40
Consensus quality: 215922 bases at least Q30
Consensus quality: 218536 bases at least Q20
Insert size: 220910; sum-of-contigs
Quality coverage: 9.03x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 15221 contig of 15221 bp in length
15322 30276 contig of 14955 bp in length
30377 40753 contig of 10377 bp in length
40854 53793 contig of 12940 bp in length
53894 65011 contig of 11118 bp in length
65112 75224 contig of 10113 bp in length
75325 84814 contig of 9490 bp in length
84915 96348 contig of 11434 bp in length
96449 106172 contig of 9724 bp in length
106273 114921 contig of 8649 bp in length
115022 123606 contig of 8585 bp in length
123707 131751 contig of 8045 bp in length
131852 138950 contig of 7099 bp in length
139051 144550 contig of 5500 bp in length
144651 152953 contig of 8303 bp in length
153054 165736 contig of 6171 bp in length
159325 165736 contig of 6412 bp in length
165837 173107 contig of 7271 bp in length
173208 178488 contig of 5281 bp in length
178589 183587 contig of 4999 bp in length
183688 187911 contig of 4224 bp in length
188012 192383 contig of 4372 bp in length
192484 196811 contig of 4328 bp in length
196912 201215 contig of 4304 bp in length
201316 205918 contig of 4603 bp in length
206019 210031 contig of 4013 bp in length
210132 213781 contig of 3650 bp in length
213882 217164 contig of 3283 bp in length
217265 219263 contig of 1999 bp in length
219364 220624 contig of 1261 bp in length
220725 221858 contig of 1134 bp in length
221959 224010 contig of 2052 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 15221: contig of 15221 bp in length
* 15222 15321: gap of 100 bp
* 15322 30276: contig of 14955 bp in length
* 30277 30376: gap of 100 bp
* 30377 40753: contig of 10377 bp in length
* 40754 40853: gap of 100 bp
* 40854 53793: contig of 12940 bp in length
* 53794 53893: gap of 100 bp
* 53894 65011: contig of 11118 bp in length
* 65012 65111: gap of 100 bp
* 65112 75224: contig of 10113 bp in length
* 75225 75324: gap of 100 bp
* 75325 84814: contig of 9490 bp in length
* 84815 84914: gap of 100 bp
* 84915 96348: contig of 11434 bp in length
* 96349 96448: gap of 100 bp
```

```
* 96449 106172: contig of 9724 bp in length
* 106173 106272: gap of 100 bp
* 114921 114921: contig of 8649 bp in length
* 114922 115021: gap of 100 bp
* 115022 123606: contig of 8585 bp in length
* 123607 123706: gap of 100 bp
* 123707 131751: contig of 8045 bp in length
* 131752 131851: gap of 100 bp
* 131852 138950: contig of 7099 bp in length
* 138951 139050: gap of 100 bp
* 139051 144550: contig of 5500 bp in length
* 144551 144650: gap of 100 bp
* 144651 152953: contig of 8303 bp in length
* 152954 153053: gap of 100 bp
* 153054 159224: contig of 6171 bp in length
* 159225 159324: gap of 100 bp
* 159325 165736: contig of 6412 bp in length
* 165737 165836: gap of 100 bp
* 165837 173107: contig of 7271 bp in length
* 173108 173207: gap of 100 bp
* 173208 178488: contig of 5281 bp in length
* 178489 178588: gap of 100 bp
* 178589 183587: contig of 4999 bp in length
* 183588 183687: gap of 100 bp
* 183688 187911: contig of 4224 bp in length
* 187912 188011: gap of 100 bp
* 188012 192383: contig of 4372 bp in length
* 192384 192483: gap of 100 bp
* 192484 196811: contig of 4328 bp in length
* 196812 196911: gap of 100 bp
* 196912 201215: contig of 4304 bp in length
* 201216 201315: gap of 100 bp
* 201316 205918: contig of 4603 bp in length
* 205919 206018: gap of 100 bp
* 206019 210031: contig of 4013 bp in length
* 210032 210131: gap of 100 bp
* 210132 213781: contig of 3650 bp in length
* 213782 213881: gap of 100 bp
* 213882 217164: contig of 3283 bp in length
* 217165 217264: gap of 100 bp
* 217265 219263: contig of 1999 bp in length
* 219264 219363: gap of 100 bp
* 219364 220624: contig of 1261 bp in length
* 220625 220724: gap of 100 bp
* 220725 221858: contig of 1134 bp in length
* 221859 221958: gap of 100 bp
* 221959 224010: contig of 2052 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..224010
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="18"
                /map="18p11.2"
                /clone="RP11-814018"
            1..15221
                /note="assembly_fragment"
            15322..30276
                /note="assembly_fragment"
            30377..40753
                /note="assembly_fragment clone_end:T7 vector_side:left"
            40854..53793
                /note="assembly_fragment"
            53894..65011
                /note="assembly_fragment"
            65112..75224
                /note="assembly_fragment"
            75325..84814
                /note="assembly_fragment"
            84915..96348
                /note="assembly_fragment"
            96449..106172
                /note="assembly_fragment"
            106273..114921
```

```
misc_feature /note="assembly_fragment"
115022..123606
misc_feature /note="assembly_fragment"
123707..131751
misc_feature /note="assembly_fragment"
131852..138950
misc_feature /note="assembly_fragment"
139051..144550
misc_feature /note="assembly_fragment clone_end:SP6 vector_side:left"
144651..152953
misc_feature /note="assembly_fragment"
153054..159224
misc_feature /note="assembly_fragment"
159325..165736
misc_feature /note="assembly_fragment"
165837..173107
misc_feature /note="assembly_fragment"
173208..178488
misc_feature /note="assembly_fragment"
178589..183587
misc_feature /note="assembly_fragment"
```

Query Match 1.0%; Score 32; DB 2; Length 224010;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 ttgaaaaaattttttttttttttttttttttttt 221
|||||

Db 165693 TTGAAAAAATTTTTTTTTTTTTTTTTTTTTT 165724
|||||

RESULT 14

PFMAL13P3/C
LOCUS PFMAL13P3 318221 bp DNA linear HTG 19-AUG-1999
DEFINITION Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL049184
VERSION AL049184.5 GI:5763803
KEYWORDS HTG; HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 318221)
AUTHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M.
and Barrett, B.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-1999) P. falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
COMMENT On Aug 24, 1999 this sequence version replaced gi:5731882.
For more information about this sequence or the Malaria Project,
see <http://www.sanger.ac.uk/Projects/P.falciparum>. IMPORTANT: This
sequence is unfinished and does not necessarily represent the
correct sequence. Work on the sequence is in progress and the
release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
Order of segments is not known; 800 n's separate segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
 source
 Location/Qualifiers
 1..318221
 /organism="Plasmodium falciparum"
 /strain="3D7"
 /db_xref="taxon:5833"
 /chromosome="13"

BASE COUNT 113610 a 29264 c 27321 g 118405 t 29621 others
ORIGIN


```
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-400N12"
/clone_lib="RPC1-11.2"
690..1132
/note="MER7C repeat: matches 173..602 of consensus"
1420..1559
/note="MER7C repeat: matches 2..173 of consensus"
1743..2156
/note="match: GSS: Em:AQ151959"
complement(3569..4072)
/note="match: GSS: Em:AQ754516"
complement(3632..4071)
/note="match: GSS: Em:AQ703632"
9776..12321
/note="L1P47 repeat: matches 3601..6141 of consensus"
12645..16537
/note="L1P47 repeat: matches 18..3601 of consensus"
16538..17237
/note="L1P10 repeat: matches 5445..6164 of consensus"
17259..17390
/note="L1MBC repeat: matches 1897..2021 of consensus"
17394..17425
/note="16 copies 2 mer aa 84% conserved"
complement(19478..20162)
/note="match: GSS: Em:AQ350792"
complement(19942..20152)
/note="match: GSS: Em:AQ806482"
19955..20086
/note="L1R29 repeat: matches 23..149 of consensus"
20236..20512
/note="L1R29 repeat: matches 370..619 of consensus"
21147..26393
/note="L1HS repeat: matches 900..6146 of consensus"
26389..27137
/note="L1HS repeat: matches 32..776 of consensus"
26723..27139
/note="Cpg island"
/evidence="not_experimental"
29280..29387
/note="54 copies 2 mer ta 60% conserved"
29459..29573
/note="23 copies 5 mer tatat 61% conserved"
29493..29590
/note="49 copies 2 mer ta 61% conserved"
29595..30568
/note="MER21B repeat: matches 14..751 of consensus"
complement(30597..30767)
/note="match: GSS: Em:AQ461439"
32936..332961
/note="13 copies 2 mer ag 92% conserved"
complement(36176..36435)
/note="match: STS: Em:L18463"
36359..36402
/note="22 copies 2 mer ct 75% conserved"
36761..36838
/note="L1P2 repeat: matches 6077..6155 of consensus"
37666..38302
/note="L1P10 repeat: matches 5526..6165 of consensus"
39311..39444
/note="67 copies 2 mer aa 60% conserved"
39632..40222
/note="L1R10C repeat: matches 1..586 of consensus"
41331..41623
/note="MER2 repeat: matches 7..308 of consensus"
complement(41701..42375)
/note="match: GSS: Em:AQ746749"
41906..42375
/note="match: GSS: Em:AQ319140"
complement(41924..42375)
/note="match: GSS: Em:A2254542"
complement(41956..42375)
/note="match: GSS: Em:AQ489422"
complement(42087..42374)
/note="match: GSS: Em:AQ808740"
complement(42095..42374)
/note="match: GSS: Em:A2517696"
complement(42105..42375)
/note="match: GSS: Em:AQ817421"
complement(42115..42393)
/note="match: GSS: Em:B94003"
42147..42408
/note="match: GSS: Em:AQ627818"
complement(42154..42374)
/note="match: GSS: Em:AQ815465"
complement(42177..42375)
/note="match: GSS: Em:AQ617736"
complement(42221..42374)
/note="match: GSS: Em:B69794"
complement(42234..42377)
/note="match: GSS: Em:AQ021494"
complement(42236..42374)
/note="match: GSS: Em:AQ037473"
42391..42718
/note="MER2 repeat: matches 327..652 of consensus"
44721..51158
/note="L1P2 repeat: matches 8..6144 of consensus"
51186..52830
/note="L1P47 repeat: matches 4480..6138 of consensus"
53104..53442
/note="match: GSS: Em:AQ376651"
complement(56787..57241)
/note="match: GSS: Em:AQ126222"
61609..62001
/note="MER61B repeat: matches 1..396 of consensus"
complement(61809..62252)
/note="match: GSS: Em:AQ429742"
63783..64278
/note="match: GSS: Em:AQ614639"
67137..67325
/note="L1R33 repeat: matches 311..519 of consensus"
71184..71219
/note="18 copies 2 mer tt 91% conserved"
73436..73677
/note="match: GSS: Em:AQ724293"
73683..73841
/note="match: GSS: Em:AQ724293"
74516..74994
/note="match: STS: Em:G63026"
75968..76017
/note="25 copies 2 mer tt 74% conserved"
76175..76272
/note="49 copies 2 mer ta 61% conserved"
78172..78283
/note="56 copies 2 mer ta 65% conserved"
complement(78272..78706)
/note="match: GSS: Em:AQ355192"
78328..78836
/note="L1PBB repeat: matches -358..155 of consensus"
78830..81169
/note="L1P2 repeat: matches 3470..5821 of consensus"
81476..81801
/note="L1P2 repeat: matches 5821..6155 of consensus"
86340..86774
/note="match: GSS: Em:AQ713295"
86358..86957
/note="match: GSS: Em:AQ426776"
86889..86925
/note="Single clone region. Sequence from clone PCR only."
88375..88444
/note="35 copies 2 mer tt 67% conserved"
complement(89320..89853)
/note="match: GSS: Em:AQ763680"
complement(89397..89870)
/note="match: GSS: Em:AQ202624"
match: STS: Em:G61256"
```


REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
2 (bases 1 to 198820)	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bada, F., Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Collangelo, M., Collins, S., Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M., Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Larcocque, K., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.	Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Feb 25, 2001 this sequence version replaced gi:7139742. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html	Direct Submission Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Feb 25, 2001 this sequence version replaced gi:7139742. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html	Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: http://www-seq.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu ----- Project Information Center project name: 333-J-10 Center clone name: 333-J-10 ----- Summary Statistics Sequencing vector: M13; M77815; 38% of reads Sequencing vector: Plasmid; N/A; 62% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.960731 Consensus quality: 197523 bases at least Q40 Consensus quality: 197946 bases at least Q30 Consensus quality: 198110 bases at least Q20 Insert size: 187000; agarose-fp Insert size: 198220; sum-of-contigs Quality coverage: 10.3 in Q20 bases; agarose-fp Quality coverage: 9.7 in Q20 * NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1	5936: contig of 5936 bp in length	5937 6036: gap of 100 bp	6037 17238: contig of 11202 bp in length	17239 17338: gap of 100 bp
17339 30763: contig of 13425 bp in length	30764 30863: gap of 100 bp	30864 50814: contig of 19951 bp in length	50815 50914: gap of 100 bp	50915 73714: contig of 22800 bp in length
73715 73814: gap of 100 bp	73815 117169: contig of 43355 bp in length	117170 117269: gap of 100 bp	117270 198820: contig of 81551 bp in length.	Location/Qualifiers 1. 198820 /organism="Homo sapiens"
FEATURES	source			

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT 54514 a 44459 c 43520 g 55727 t 600 others

ORIGIN

Query Match 0.9%; Score 31; DB 2; Length 198820;
 Best Local Similarity 100.0%; Pred. No. 0.00052;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 aaaaaattttttttttttttttttttttttttttgt 223
 |||
 Db 18920 AAAAAAAAAATTTTATTTTATTTTATTTTATTTTGT 18890

RESULT 20

AL671090/LOCUS AC079869 198932 bp DNA linear HTG 08-FEB-2002

DEFINITION Mus musculus chromosome 4 clone RP23-154A10, *** SEQUENCING IN PROGRESS ***, in unordered pieces.

ACCESSION AL671090

VERSION AL671090.4 GI:18643845

KEYWORDS HTG; HTGS-PHASE1.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (sites)

AUTHORS Plumb,B.

TITLE Direct Submission

JOURNAL Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Feb 10, 2002 this sequence version replaced gi.18491462.

COMMENT ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BM154A10
 ----- Summary Statistics
 Sequencing program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 95% of reads
 Chemistry: Dye-terminator; 4% of reads
 Consensus quality: 197921 bases at least Q40
 Consensus quality: 198121 bases at least Q30
 Consensus quality: 198350 bases at least Q20
 Insert size: 198732; sum-of-contigs
 Insert size: 194340; 2.7% error; agarose-fp
 Quality coverage: 10.28x in Q20 bases; sum-of-contigs Quality coverage: 10.51x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source

1..198932
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="4"
 /clone="RP23-154A10"
 /clone_lib="RPCI-23"
 1..65901
 /note="assembly_fragment:00047"
 66002..192023
 /note="assembly_fragment:02231
 fragment_chain:1"
 192124..198932
 /note="assembly_fragment:01866
 fragment_chain:1
 clone_end:T7
 vector_side:right"
 66075 a 35350 c 35808 g 61499 t 200 others

BASE COUNT 66075 a 35350 c 35808 g 61499 t 200 others

ORIGIN

Query Match 0.9%; Score 31; DB 2; Length 198932;
 Best Local Similarity 100.0%; Pred. No. 0.00052;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1519 gagaagagagagagagagagagagagagag 1549
 |||
 Db 24229 GAGAGGAGAGAGGAGAGAGAGAGAGAGG 24199

RESULT 21

AC079869/LOCUS AC079869 243675 bp DNA linear ROD 11-JUL-2001

DEFINITION Mus musculus 9 BAC RP23-49M18 (Roswell Park Cancer Institute Mouse BAC Library) complete sequence.

ACCESSION AC079869

VERSION AC079869.22 GI:14670076

KEYWORDS HTG.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 243675)

AUTHORS Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C., Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T., Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J., Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z., Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R., Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M., Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S., Kovar,C., Liu,J., Liu,W., Loulsegad,H., Lozado,R.J., Martin,R., Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S., Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B., Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S., Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A., Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G., Worley,K. and Gibbs,R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 243675)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (15-SEP-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 243675)

AUTHORS Worley,K.C.

TITLE Direct Submission


```

repeat_region      /rpt_family="B2"
complement(29457..29604)
repeat_region      /rpt_family="B1-F"
complement(30224..30344)
repeat_region      /rpt_family="Pb1"
complement(30594..30753)
repeat_region      /rpt_family="RSINEL"
30762..30785
repeat_region      /rpt_family="AT_rich"
31123..31147
repeat_region      /rpt_family="(TG)n"
complement(31302..31433)
repeat_region      /rpt_family="B1-F"
31444..31503
repeat_region      /rpt_family="(TA)n"
complement(31553..31665)
repeat_region      /rpt_family="RSINEL"

```

Query Match 0.9%; Score 31; DB 10; Length 243675;
 Best Local Similarity 100.0%; Pred. No. 0.00051;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 191 tgaaaaaatTTTTTTTTTTTTTTTTTTT 221
Db 137978 TGAAAAAATTTTTTTTTTTTTTTTTT 137948

```

RESULT 22

```

LOCUS      AX277981      8801 bp      DNA      linear      PAT 01-NOV-2001
DEFINITION Sequence 144 from Patent WO0177375.
ACCESSION  AX277981
VERSION     AX277981.1 GI:16605053
KEYWORDS   .
SOURCE      synthetic construct.
ORGANISM    artificial sequence.
REFERENCE   1 (sites)
AUTHORS     Olek, A., Piepenbrock, C. and Berlin, K.
TITLE       Diagnosis of diseases associated with gene regulation
JOURNAL     Patent: WO 0177375-A 144 18-OCT-2001;
Epigenomics AG (DE)
FEATURES    Location/Qualifiers
            source
            1..8801
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="chemically treated genomic DNA (Homo sapiens)"

```

```

BASE COUNT  2344 a 366 c 2057 g 4034 t
ORIGIN

```

Query Match 0.9%; Score 30; DB 6; Length 8801;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 192 gaaaaaatTTTTTTTTTTTTTTTTTTT 221
Db 971 GAAAAAATTTTTTTTTTTTTTTTTT 1000

```

RESULT 23

```

LOCUS      AX323672      8801 bp      DNA      linear      PAT 07-JAN-2002
DEFINITION Sequence 160 from Patent WO0192565.
ACCESSION  AX323672
VERSION     AX323672.1 GI:18094420
KEYWORDS   .
SOURCE      synthetic construct.
ORGANISM    artificial sequence.
REFERENCE   1 (sites)
AUTHORS     Olek, A., Piepenbrock, C. and Berlin, K.
TITLE       Diagnosis of diseases associated with dna transcription

```

```

JOURNAL     Patent: WO 0192565-A 160 06-DEC-2001;
Epigenomics AG (DE)
FEATURES    Location/Qualifiers
            source
            1..8801
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="chemically treated genomic DNA (Homo sapiens)"

```

```

BASE COUNT  2344 a 366 c 2057 g 4034 t
ORIGIN

```

Query Match 0.9%; Score 30; DB 6; Length 8801;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 192 gaaaaaatTTTTTTTTTTTTTTTTTTT 221
Db 971 GAAAAAATTTTTTTTTTTTTTTTTT 1000

```

RESULT 24

```

LOCUS      AX346643      8801 bp      DNA      linear      PAT 01-FEB-2002
DEFINITION Sequence 1714 from Patent WO0200928.
ACCESSION  AX346643
VERSION     AX346643.1 GI:18494529
KEYWORDS   .
SOURCE      synthetic construct.
ORGANISM    synthetic construct.
REFERENCE   1 (sites)
AUTHORS     Olek, A., Piepenbrock, C. and Berlin, K.
TITLE       Diagnosis of diseases associated with the immune system
JOURNAL     Patent: WO 0200928-A 1714 03-JAN-2002;
Epigenomics AG (DE)
FEATURES    Location/Qualifiers
            source
            1..8801
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="chemically treated genomic DNA (Homo sapiens)"

```

```

BASE COUNT  2344 a 366 c 2057 g 4034 t
ORIGIN

```

Query Match 0.9%; Score 30; DB 6; Length 8801;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 192 gaaaaaatTTTTTTTTTTTTTTTTTTT 221
Db 971 GAAAAAATTTTTTTTTTTTTTTTTT 1000

```

RESULT 25

```

LOCUS      HSA011712/c  16689 bp      DNA      linear      PRI 27-APR-1999
DEFINITION Homo sapiens TNNT1 gene, exons 1-11 (and joined CDS).
ACCESSION  AJ011712
VERSION     AJ011712.1 GI:4056561
KEYWORDS   slow skeletal; TNNT1 gene; troponin T.
SOURCE      human
ORGANISM    Homo sapiens

```

```

REFERENCE   1 (bases 1 to 16689)
AUTHORS     Barton, P.J., Cullen, M.E., Townsend, P.J., Brand, N.J., Mullen, A.J.,
Norman, D.A., Bhavsar, P.K. and Yacoub, M.H.
TITLE       Close physical linkage of human troponin genes: organization,
sequence, and expression of the locus encoding cardiac troponin I
and slow skeletal troponin T
JOURNAL     Genomics 57 (1), 102-109 (1999)
MEDLINE     99208666
REFERENCE   2 (bases 1 to 16689)

```

```

AUTHORS      Barton, P.J.R.
TITLE        Direct Submission
JOURNAL      Submitted (15-SEP-1998) Barton P.J.R., Cardiothoracic Surgery,
              National Heart & Lung Institute, Dovehouse Street, London SW3 6JB,
              United Kingdom
REMARK       Revised on 20-DEC-1998
COMMENT      Related sequences : X90780 (cardiac troponin I gene, directly
              upstream)
FEATURES
  source      AJ011713 (TNNT1 gene, exons 12-14).
               Location/Qualifiers
               1..16889
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="19"
               /map="q13.4"
               /clone="DMPC-HFF#1-1495-E10"
  promoter    1..2580
               /gene="TNNT1"
               /note="promoter region"
  gene         join(1..16889,AJ011713.1:1..1846)
               /gene="TNNT1"
               2552..2557
  TATA_signal  /gene="TNNT1"
               /note="putative TATA box"
  prim_transcript join(2581..16889,AJ011713.1:1..1846)
               /gene="TNNT1"
  exon         2581..2626
               /gene="TNNT1"
               /number=1
  mRNA         join(2581..2626,4587..4629,4734..4747,5048..5074,
               13679..5321,6189..6210,9829..9892,10447..10563,10789..10866,
               AJ011713.1:471..609,AJ011713.1:748..788,
               AJ011713.1:1715..1846)
               /gene="TNNT1"
               /note="alternatively spliced mRNA"
               /product="troponin T, slow skeletal muscle"
  mRNA         join(2581..2626,4587..4629,4734..4747,5048..5074,
               5289..5321,6189..6210,9829..9892,10447..10563,
               10789..10866,13679..13792,14577..14686,
               AJ011713.1:519..609,AJ011713.1:748..788,
               AJ011713.1:1715..1846)
               /gene="TNNT1"
               /note="alternatively spliced mRNA"
               /product="troponin T, skeletal muscle"
  mRNA         join(2581..2626,4587..4629,4734..4747,5048..5074,
               5289..5321,6189..6210,9829..9892,10447..10563,
               10789..10866,13679..13792,14577..14686,
               AJ011713.1:519..609,AJ011713.1:748..788,
               AJ011713.1:1715..1846)
               /gene="TNNT1"
               /note="alternatively spliced mRNA"
               /product="troponin T, skeletal and cardiac muscle"
  5'UTR        join(2581..2626,4587..4597)
               /gene="TNNT1"
               4587..4629
               /gene="TNNT1"
               /number=2
  CDS          join(4598..4629,4734..4747,5048..5074,5289..5321,
               6189..6210,9829..9892,10447..10563,10789..10866,
               13679..13792,14577..14686,AJ011713.1:471..609,
               AJ011713.1:748..788,AJ011713.1:1715..1760)
               /gene="TNNT1"
               /note="alternatively spliced isoform"
               /codon_start=1
               /product="slow skeletal muscle troponin T"
               /protein_id="CAA09752.1"
               /db_xref="GI:4056564"
               /translation="MSDTEQEYEEQPEEAEAEAEPEEPVPAEPEERPKPSR
               PVVPLIPPKIPEGERVDFDDIHRKMEKDLLEQLTLDVHFQKKEEELVALKER
               TERRSRAEQRTKEKREQAALAEKMKKEEAKKRAEDDAKKKVLNNGAH
               FGGYLVKAEQKRGKQRTGRENKVRILSERKKPLDIDYNGEQLREKAQELSDWIHQLE
               ARAKAQELSDWIHQLESEKFDLMKAKQKQKYEINVLNRIASHQKFRGAGKRGVGR
               WK"
  CDS          join(4598..4629,4734..4747,5048..5074,5289..5321,
               6189..6210,9829..9892,10447..10563,10789..10866,
               13679..13792,14577..14686,AJ011713.1:519..609,
               AJ011713.1:748..788,AJ011713.1:1715..1760)
               /gene="TNNT1"
               /note="alternatively spliced isoform"
               /codon_start=1
               /product="slow skeletal muscle troponin T"
               /protein_id="CAA09751.1"
               /db_xref="GI:4056563"
               /translation="MSDTEQEYEEQPEEAEAEAEPEEPVPAEPEERPKPSR
               PVVPLIPPKIPEGERVDFDDIHRKMEKDLLEQLTLDVHFQKKEEELVALKER
               TERRSRAEQRTKEKREQAALAEKMKKEEAKKRAEDDAKKKVLNNGAH
               FGGYLVKAEQKRGKQRTGRENKVRILSERKKPLDIDYNGEQLREKAQELSDWIHQLE
               SEKFDLMKAKQKQKYEINVLNRIASHQKFRGAGKRGVGRWK"
  CDS          join(4598..4629,4734..4747,5048..5074,6189..6210,
               9829..9892,10447..10563,10789..10866,13679..13792,
               14577..14686,AJ011713.1:519..609,AJ011713.1:748..788,
               AJ011713.1:1715..1760)
               /gene="TNNT1"
               /note="alternatively spliced isoform"
               /codon_start=1
               /product="slow skeletal muscle troponin T"
               /protein_id="CAA09750.1"
               /db_xref="GI:4056562"
               /translation="MSDTEQEYEEQPEEAEAEAEPEERPKPSRVVPLIPPKI
               PEGERVDFDDIHRKMEKDLLEQLTLDVHFQKKEEELVALKERTERRSRAEQ
               ORFTEKEREQAALAEKMKKEEAKKRAEDDAKKKVLNNGAHFGGYLVKAEQK
               RGRKQRTGRENKVRILSERKKPLDIDYNGEQLREKAQELSDWIHQLESEKFDLMKAK
               LK
               QCKYKYEINVLNRIASHQKFRGAGKRGVGRWK"
               4734..4747
               /gene="TNNT1"
               /number=3
               5048..5074
               /gene="TNNT1"
               /number=4
               5289..5321
               /gene="TNNT1"
               /note="alternative exon"
               /number=5
               6189..6210
               /gene="TNNT1"
               /number=6
               9829..9892
               /gene="TNNT1"
               /number=7
               10447..10563
               /gene="TNNT1"
               /number=8
               10789..10866
               /gene="TNNT1"
               /number=9
               13679..13792
               /gene="TNNT1"
               /number=10
               14577..14686
               /gene="TNNT1"
               /number=11
               14687..>16689
               /gene="TNNT1"
               /note="approx. 200 bp gap with AJ011713"
               /number=11
  BASE COUNT   3888 a 4422 g 4743 g 3615 t 21 others
  ORIGIN
  Query Match      0.9%; Score 30; DB 9; Length 16689;
  Best Local Similarity 100.0%; Pred. No. 0.0021;
  Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  Qy 1526 aqaaggggaagagagagagagagaggg 1555
  DB 6448 AGAGGGGAAGGAGGAGGAGGAGG 6419

```



```

Qy 193 aaaaaattttttttttttttttttttttttt 222
|||||ttttttttttttttttttttttttttttttt
Db 15493 AAAAAAATTTTttttttttttttttttttt 15464

RESULT 27
LOCUS AC015033/c
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC015033
VERSION AC015033.1 GI:6436302
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 58569)
Adams,M. and Venter,J.C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10213386 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source Location/Qualifiers
1..58569
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 16385 a 12617 c 12623 g 16944 t
ORIGIN

Query Match 0.9%; Score 30; DB 2; Length 58569;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 195 aaaaaattttttttttttttttttttttttt 224
|||||ttttttttttttttttttttttttttttttt
Db 53307 AAAAAATTTTttttttttttttttttttt 53278

RESULT 28
LOCUS AC024343
DEFINITION Homo sapiens chromosome 8 clone RP11-380N6 map 8, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC024343
VERSION AC024343.1 GI:7108120
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 64255)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-380N6
Unpublished
2 (bases 1 to 64255)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhaltier,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepl,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,

```

```

Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivares,T.M.,
Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5526
Center clone name: 380_N_L6
-----
* NOTE: This record contains 72 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
765: contig of 765 bp in length
766 865: gap of 100 bp
866 1681: contig of 816 bp in length
1682 1781: gap of 100 bp
1782 2592: contig of 811 bp in length
2593 2692: gap of 100 bp
2693 3452: contig of 760 bp in length
3453 3552: gap of 100 bp
3553 4336: contig of 784 bp in length
4337 4436: gap of 100 bp
4437 5217: contig of 781 bp in length
5218 5317: gap of 100 bp
5318 6111: contig of 794 bp in length
6112 6211: gap of 100 bp
6212 6995: contig of 784 bp in length
6996 7095: gap of 100 bp
7096 7885: contig of 790 bp in length
7886 7985: gap of 100 bp
7986 8767: contig of 782 bp in length
8768 8867: gap of 100 bp
8868 9669: contig of 802 bp in length
9670 9769: gap of 100 bp
9770 10578: contig of 809 bp in length
10579 10678: gap of 100 bp
10679 11482: contig of 804 bp in length
11483 11582: gap of 100 bp
11583 12364: contig of 782 bp in length
12365 12464: gap of 100 bp
12465 13256: contig of 792 bp in length
13257 13356: gap of 100 bp
13357 14138: contig of 782 bp in length
14139 14238: gap of 100 bp
14239 15045: contig of 807 bp in length
15046 15145: gap of 100 bp
15146 15964: contig of 819 bp in length
15965 16064: gap of 100 bp

```

TITLE
JOURNAL

COMMENT

Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L13222

Center Clone name: 960_P_18

* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 720: contig of 720 bp in length
* 721 820: gap of 100 bp
* 821 1575: contig of 755 bp in length
* 1576 1675: gap of 100 bp
* 1676 2407: contig of 732 bp in length
* 2408 2507: gap of 100 bp
* 2508 3249: contig of 742 bp in length
* 3250 3349: gap of 100 bp
* 3350 4100: contig of 751 bp in length
* 4101 4200: gap of 100 bp
* 4201 4948: contig of 748 bp in length
* 4949 5048: gap of 100 bp
* 5049 5792: contig of 744 bp in length
* 5793 5892: gap of 100 bp
* 5893 6629: contig of 737 bp in length
* 6630 6729: gap of 100 bp
* 6730 7494: contig of 765 bp in length
* 7495 7594: gap of 100 bp
* 7595 8292: contig of 698 bp in length
* 8293 8392: gap of 100 bp
* 8393 9137: contig of 745 bp in length
* 9138 9237: gap of 100 bp
* 9238 9959: contig of 722 bp in length
* 9960 10059: gap of 100 bp
* 10060 10806: contig of 747 bp in length
* 10807 10906: gap of 100 bp
* 10907 11649: contig of 743 bp in length
* 11650 11749: gap of 100 bp
* 11750 12460: contig of 711 bp in length
* 12461 12560: gap of 100 bp
* 12561 13324: contig of 764 bp in length
* 13325 13424: gap of 100 bp
* 13425 14169: contig of 745 bp in length
* 14170 14269: gap of 100 bp
* 14270 15043: contig of 774 bp in length
* 15044 15143: gap of 100 bp
* 15144 15919: contig of 776 bp in length
* 15920 16019: gap of 100 bp
* 16020 16737: contig of 718 bp in length
* 16738 16837: gap of 100 bp
* 16838 17518: contig of 681 bp in length
* 17519 17618: gap of 100 bp
* 17619 18341: contig of 723 bp in length

* 18342 18441: gap of 100 bp
* 18442 19167: contig of 726 bp in length
* 19168 19267: gap of 100 bp
* 19268 19999: contig of 732 bp in length
* 20000 20099: gap of 100 bp
* 20100 20835: contig of 736 bp in length
* 20836 20935: gap of 100 bp
* 20936 21690: contig of 755 bp in length
* 21691 21790: gap of 100 bp
* 21791 22527: contig of 737 bp in length
* 22528 22627: gap of 100 bp
* 22628 23387: contig of 760 bp in length
* 23388 23487: gap of 100 bp
* 23488 24234: contig of 747 bp in length
* 24235 24334: gap of 100 bp
* 24335 25086: contig of 752 bp in length
* 25087 25186: gap of 100 bp
* 25187 25934: contig of 748 bp in length
* 25935 26034: gap of 100 bp
* 26035 26757: contig of 723 bp in length
* 26758 26857: gap of 100 bp
* 26858 27595: contig of 738 bp in length
* 27596 27695: gap of 100 bp
* 27696 28428: contig of 733 bp in length
* 28429 28528: gap of 100 bp
* 28529 29287: contig of 759 bp in length
* 29288 29387: gap of 100 bp
* 29388 30133: contig of 746 bp in length
* 30134 30233: gap of 100 bp
* 30234 30954: contig of 721 bp in length
* 30955 31054: gap of 100 bp
* 31055 31795: contig of 741 bp in length
* 31796 31895: gap of 100 bp
* 31896 32644: contig of 749 bp in length
* 32645 32744: gap of 100 bp
* 32745 33516: contig of 772 bp in length
* 33517 33616: gap of 100 bp
* 33617 34369: contig of 753 bp in length
* 34370 34469: gap of 100 bp
* 34470 35200: contig of 731 bp in length
* 35201 35300: gap of 100 bp
* 35301 36048: contig of 748 bp in length
* 36049 36148: gap of 100 bp
* 36149 36894: contig of 746 bp in length
* 36895 36994: gap of 100 bp
* 36995 37729: contig of 735 bp in length
* 37730 37829: gap of 100 bp
* 37830 38565: contig of 736 bp in length
* 38566 38665: gap of 100 bp
* 38666 39412: contig of 747 bp in length
* 39413 39512: gap of 100 bp
* 39513 40273: contig of 761 bp in length
* 40274 40373: gap of 100 bp
* 40374 41112: contig of 739 bp in length
* 41113 41212: gap of 100 bp
* 41213 41969: contig of 757 bp in length
* 41970 42069: gap of 100 bp
* 42070 42819: contig of 750 bp in length
* 42820 42919: gap of 100 bp
* 42920 43671: contig of 752 bp in length
* 43672 43771: gap of 100 bp
* 43772 44509: contig of 738 bp in length
* 44510 44609: gap of 100 bp
* 44610 45346: contig of 737 bp in length
* 45347 45446: gap of 100 bp
* 45447 46213: contig of 767 bp in length
* 46214 46313: gap of 100 bp
* 46314 47076: contig of 763 bp in length
* 47077 47176: gap of 100 bp
* 47177 47929: contig of 753 bp in length
* 47930 48029: gap of 100 bp
* 48030 48784: contig of 755 bp in length
* 48785 48884: gap of 100 bp

```

* 48885 49635: contig of 751 bp in length
* 49636 49735: gap of 100 bp
* 49736 50474: contig of 739 bp in length
* 50475 50574: gap of 100 bp
* 50575 51317: contig of 743 bp in length
* 51318 51417: gap of 100 bp
* 51418 52169: contig of 752 bp in length
* 52170 52269: gap of 100 bp
* 52270 53008: contig of 739 bp in length
* 53009 53108: gap of 100 bp
* 53109 53861: contig of 753 bp in length
* 53862 53961: gap of 100 bp
* 53962 54724: contig of 763 bp in length
* 54725 54824: gap of 100 bp
* 54825 55571: contig of 747 bp in length
* 55572 55671: gap of 100 bp
* 55672 56448: contig of 777 bp in length
* 56449 56548: gap of 100 bp
* 56549 57289: contig of 741 bp in length
* 57290 57389: gap of 100 bp
* 57390 58100: contig of 711 bp in length
* 58101 58200: gap of 100 bp

```

```

Query Match          0.9%: Score 30; DB 2; Length 67462;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 193 aaaaaattttttttttttttttttttttttt 222
|||||
Db 21115 AAAAAATTTTTTTTTTTTTTTTTTTTG 21086

```

```

RESULT 30
AC109316
LOCUS Homo sapiens chromosome 8 clone RP11-96H13 map 8, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
AC109316
VERSION AC109316.1 GI:18482245
KEYWORDS .HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 73858)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-96H13
Unpublished
2 (bases 1 to 73858)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye.W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome

```

COMMENT

```

Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23111
Center clone name: 96_H13
-----

```

```

* NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

```

* 1 713: contig of 713 bp in length
* 714 813: gap of 100 bp
* 814 1531: contig of 718 bp in length
* 1532 1631: gap of 100 bp
* 1632 2360: contig of 729 bp in length
* 2361 2460: gap of 100 bp
* 2461 3148: contig of 688 bp in length
* 3149 3248: gap of 100 bp
* 3249 3939: contig of 691 bp in length
* 3940 4039: gap of 100 bp
* 4040 4757: contig of 718 bp in length
* 4758 4857: gap of 100 bp
* 4858 5576: contig of 719 bp in length
* 5577 5676: gap of 100 bp
* 5677 6390: contig of 714 bp in length
* 6391 6490: gap of 100 bp
* 6491 7212: contig of 722 bp in length
* 7213 7312: gap of 100 bp
* 7313 8025: contig of 713 bp in length
* 8026 8125: gap of 100 bp
* 8126 8865: contig of 740 bp in length
* 8866 8965: gap of 100 bp
* 8966 9679: contig of 714 bp in length
* 9680 9779: gap of 100 bp
* 9780 10518: contig of 739 bp in length
* 10519 10618: gap of 100 bp
* 10619 11349: contig of 731 bp in length
* 11350 11449: gap of 100 bp
* 11450 12174: contig of 725 bp in length
* 12175 12274: gap of 100 bp
* 12275 12987: contig of 713 bp in length
* 12988 13087: gap of 100 bp
* 13088 13779: contig of 692 bp in length
* 13780 13879: gap of 100 bp
* 13880 14593: contig of 714 bp in length
* 14594 14693: gap of 100 bp
* 14694 15420: contig of 727 bp in length
* 15421 15520: gap of 100 bp
* 15521 16256: contig of 736 bp in length
* 16257 16356: gap of 100 bp
* 16357 17093: contig of 737 bp in length
* 17094 17193: gap of 100 bp
* 17194 17920: contig of 727 bp in length
* 17921 18020: gap of 100 bp
* 18021 18730: contig of 710 bp in length
* 18731 18830: gap of 100 bp
* 18831 19560: contig of 730 bp in length
* 19561 19660: gap of 100 bp
* 19661 20394: contig of 734 bp in length

```

50078	50767:	contig of 690 bp	in length
50768	50867:	gap of 100 bp	
50868	51579:	contig of 712 bp	in length
51580	51679:	gap of 100 bp	
51680	52398:	contig of 717 bp	in length
52397	52496:	gap of 100 bp	
52497	53229:	contig of 733 bp	in length
53230	53329:	gap of 100 bp	
53330	54061:	contig of 732 bp	in length
54062	54161:	gap of 100 bp	
54162	54880:	contig of 719 bp	in length
54881	54980:	gap of 100 bp	
54981	55690:	contig of 710 bp	in length
55691	55790:	gap of 100 bp	
55791	56522:	contig of 737 bp	in length

	Qy	192	gaaaaaaattttttttttttttttttt 221
	Dd	27325	GAAAAAAATTTTTTTTTTTTTTTT 27354

Search completed: June 14, 2002, 22:06:41
Job time: 23350 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 17:02:26 ; Search time 72.29 Seconds
(without alignments)
11338.761 Million cell updates/sec

Title: US-09-667-130-1

Perfect score: 3337

Sequence: 1 gaattccggtaaagtaacaa.....tcttataaaataaataattc 3337

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3337	100.0	3337	1 US-08-072-610-1	Sequence 1, Appli
2	3337	100.0	3337	2 US-08-719-822B-1	Sequence 1, Appli
3	3337	100.0	3337	4 US-09-092-458-1	Sequence 1, Appli
4	29	0.9	32	5 PCT-US92-10792-1	Sequence 16, Appli
5	28	0.8	90	4 US-09-065-058-16	Sequence 7, Appli
6	27	0.8	50	2 US-08-828-007-7	Sequence 4, Appli
7	27	0.8	3138	1 US-07-867-106-4	Sequence 1, Appli
8	26	0.8	76	4 US-09-244-794A-1	Sequence 1, Appli
9	26	0.8	76	4 US-09-247-190-1	Sequence 1, Appli
10	26	0.8	596	1 US-08-150-331-45	Sequence 45, Appli
11	26	0.8	1380	2 US-08-467-559B-1	Sequence 1, Appli
12	26	0.8	2406	1 US-08-396-479B-7	Sequence 7, Appli
13	26	0.8	2406	1 US-08-818-823-7	Sequence 7, Appli
14	26	0.8	19736	4 US-09-740-035-3	Sequence 3, Appli
15	25	0.7	337	4 US-09-328-111-586	Sequence 586, App
16	25	0.7	342	4 US-09-385-982-342	Sequence 342, App
17	25	0.7	523	2 US-08-628-413-1	Sequence 1, Appli
18	25	0.7	547	1 US-08-131-365B-41	Sequence 41, Appli
19	25	0.7	547	2 US-08-668-123-41	Sequence 41, Appli
20	25	0.7	567	4 US-09-385-982-427	Sequence 427, App
21	25	0.7	587	1 US-08-313-681A-3	Sequence 3, Appli
22	25	0.7	593	3 US-09-322-911-3	Sequence 3, Appli
23	25	0.7	607	4 US-09-385-982-262	Sequence 262, App
24	25	0.7	629	4 US-09-385-982-288	Sequence 288, App
25	25	0.7	629	4 US-09-385-982-389	Sequence 389, App
26	25	0.7	730	4 US-09-352-990-3	Sequence 3, Appli
27	25	0.7	770	3 US-08-865-297-5	Sequence 5, Appli

28	25	0.7	777	4 US-09-385-982-531	Sequence 531, App
29	25	0.7	809	1 US-08-441-629-1	Sequence 1, Appli
30	25	0.7	809	3 US-08-776-207-1	Sequence 1, Appli
31	25	0.7	809	5 PCT-US95-09172-1	Sequence 1, Appli
32	25	0.7	1225	1 US-08-286-020-1	Sequence 1, Appli
33	25	0.7	1225	1 US-08-603-919-1	Sequence 1, Appli
34	25	0.7	1227	1 US-08-080-386-1	Sequence 1, Appli
35	25	0.7	1227	2 US-08-390-000A-1	Sequence 1, Appli
36	25	0.7	1621	4 US-09-020-956-107	Sequence 107, App
37	25	0.7	1621	4 US-09-030-607-107	Sequence 107, App
38	25	0.7	1621	4 US-09-439-313-107	Sequence 107, App
39	25	0.7	1683	4 US-09-347-803-11	Sequence 11, Appli
40	25	0.7	1690	4 US-09-071-224-1	Sequence 1, Appli
41	25	0.7	1813	4 US-09-071-224-3	Sequence 3, Appli
42	25	0.7	1889	3 US-09-187-050-1	Sequence 1, Appli
43	25	0.7	1921	2 US-08-557-128-11	Sequence 11, Appli
44	25	0.7	1964	1 US-08-132-168A-31	Sequence 31, Appli
45	25	0.7	1968	2 US-08-937-540-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-072-610-1
; Sequence 1, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,610
; FILING DATE: 19930602
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/07686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pVMB3.3.1
US-08-072-610-1

Query Match

100.0%; Score 3337; DB 1; Length 3337;

Best Local Similarity 100.0%; Pred. No. 0;				
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	gaattccggtaaaagtaaacactatggttcctgctatctatataaaaccttactatttctac	60	
DB	1	GAATTCGGGTAAAGTAACAACTATGCTTTCGATCTATATATATAACCTTACTAATTTTATC	60	
QY	61	ttttgcttttcttttaatttcctcctcaacagtaagataaaaaataatctataaaactgc	120	
DB	61	TTTTTGCTTTCTTTTAATTCATGCTTCAACAGTAAGATAAAAAATAATCTATAAAAACTGC	120	
QY	121	tatatatacatatattcatatgaatggcatttggtaattgcgatcatcttaatttacgta	180	
DB	121	TATATATACATATATATTCATAGTGGCATTTGTGAATTCGGATCATTTTAAATTTACGTA	180	
QY	181	aaacaataatgaaaaaatt	240	
DB	181	AAACAATAATTGAAAAAATTT	240	
QY	241	aattgaaaatgctctgatgattgttagagtgagagatccttcaaacgcggttttag	300	
DB	241	AATTGCAAAATGCTTCTGATGATGTTGTAGAGTGGAGGATCCTTCAACGACGGTTTAG	300	
QY	301	aattagaagagaaaattttgatgagaattcaggtgatgatgaacctcttttagatgcta	360	
DB	301	AATTAGAAGAGCAAAATTTTGTATGAGAAATTCAGGTGATGATGAAACTCTTTTAGATGCTA	360	
QY	361	ccccgaagatgacttgccttaacagatttccaaatgaacacgatgaggaagcaacg	420	
DB	361	CCCCGAAGATGACTTTGGCTTTAACAGATTGCCAATTTGAACAGATGAAGAACTCAACG	420	
QY	421	aaacgttagatgagtgtaacattagaggaggtttccactgaagatatggaacagag	480	
DB	421	AAACGTTAGATGAGGTGAATCATTTAGGAGGTTTCCACTCAAGATATGGAACAGAG	480	
QY	481	atggtcacaagatgatacggaacagaagaaggactacctggtgatattggaaggagaag	540	
DB	481	ATGGCTCAACAGATGATACGGAACAGAAAGAGAGGACTACTCTGTTGATATGGAAGGAGAAG	540	
QY	541	aaagaactggcgatattggaagcaggggaagaagctgggtatttggaaagcaggggaagaaa	600	
DB	541	AAGAAGCTGGCGATATGGAAGCAGGGGAAGAAGCTGGTCATTTGGNAGCAGGGGAAGAAA	600	
QY	601	ctggcgatttgaagcaggggaagaactggcgatttggaaagcaggggaagaagctggcg	660	
DB	601	CTGGCGATTTGGAAGCAGGGGAAGAACTGGCGATTTGGAAGCAGGGGAAGAAGCTGGTG	660	
QY	661	atttgaagcaggggaagaactggcgatttggaaagcaggggaagaagaactggagatgcg	720	
DB	661	ATTTGGAAGCAGGGGAAGAAGAACTGGCGATTTGGAAGCAGGGGAAGAAGCTGGAGATGCGG	720	
QY	721	aaactgaagaagagcaactggagatgcggaactgaaatggagcaactgtgtatgttag	780	
DB	721	AAACTGAAGAAGAGCAACTGGAGATGCGGAACCTGAATGGACCAACTGTGTATGTAG	780	
QY	781	acacagaagatagttcagctgatgagcagaagaagatcacatgtcctgcctcaagaaaaag	840	
DB	781	ACACAGAAGATAGTTCAGCTGATGGAGCAGAAAAGTACATGTCTCTGCTCAAGAAAATG	840	
QY	841	tacaacctcccatagatgacccctcttggaaagtatttggataaagataaattt	900	
DB	841	TACAACCTCCGATAGTATGATGCCCCTCTTTGGAAGTATTTTGGATAAAGATATAATTT	900	
QY	901	ttgatcatataaagatttcagaccactatctgaacaaatttggcggtactgctcaaac	960	
DB	901	TTGATCATATTAAGATTTTCAGCCCACTATTTCGAACAAATTTGGCGGGTACTGCTTAAC	960	
QY	961	atgttcagggaagaatttgcgaatgaacctgtaccattacacagtggcagaagaccgcg	1020	
DB	961	ATGTTACGGGACAAAGAAATTTGCCAATGAAACCTGTACCATTTACACGTGGCAGAGGCCCG	1020	
QY	1021	cqcaagtaccacgggaagattagatgccactccagagatgacttcgattagatgttta	1080	
DB	1021	TT	1080	

DB	1021	CCCAAGTACCACGGGAAGAAATTTAGATGCCACTCCAGAGGTGACTTCGATTTAGATGTTA	1080	
QY	1081	cagaatctcccagaggaagtagaattagtagatgaagagcaactgaagaagaatcaa	1140	
DB	1081	CAGAACTCTCCGAGGAGTAGAATTTAGTATTAGATGAAGAGCAACTGAAGAAGAAATCAA	1140	
QY	1141	cggaaagtgggaccacgaagaaggaacccaaacccgaagaatttagatgccactccagagatg	1200	
DB	1141	CGAAAGTGGGACCAACGGGAAGAACCCAGAAATTCAGATGCCACTCCAGAGGATG	1200	
QY	1201	gatttcgcattagacaaactcgaagaagagaacagaagaacatagaggagagaagaa	1260	
DB	1201	GATTTTCGATTTAGACAAACTGCAGAAGGAGAAACAGAAACCTAGAGGGAGAGAAA	1260	
QY	1261	cagaagaagctcgaagaagagatcagaagaactccagaaggaaggaagagattag	1320	
DB	1261	CAGAAGAGCTGCAGAAGGAGAAATTCAGAACTCCAGAAGGAGAGAGATTAG	1320	
QY	1321	aggcaactccagaggatgatttcgcattagatggaactacattagaagaacccgaagaaa	1380	
DB	1321	AGGCAACTCCAGAGGATGATTTCCGATTTAGATGGAACCTACATTAGAAGAAACCGAAGAA	1380	
QY	1381	ctgcagaagagagaagaacccgtagaggagaagaacccgtagaggagaagaacccgtag	1440	
DB	1381	CTGCAGAAGGAGAAAACCCGTAGAGGGAGAAACCCGTAGGGGAGAGAAACCCGTAG	1440	
QY	1441	agggagaagaagctcgaagaagagagagaggttagaggaactccagagagatgacttc	1500	
DB	1441	AGGGAGAGAAGCTCGAGNAGGAGAGAGAGTTAGAGGCAACTCCAGAGGATGACTTCC	1500	
QY	1501	aattagaagaacccatcagggaggaagggggaaggggaagggggaaggggaaggaag	1560	
DB	1501	AATTAGAAGAACCATCAGGAGAAGGAGAAGGGGAAGAGAGAGGAGGAGGAGAGAG	1560	
QY	1561	gagaagcgttagtagcagtgccagtagtggccgaacccgtagaactagtagtactccttc	1620	
DB	1561	GAGAAGCCTTTAGTAGCAGTGGCAGTAGTGGCCGAAACCCGTAGAGTAGTACTCTCTGCTC	1620	
QY	1621	agcctgtcaaaccaatggtcgctccaaacgagagtgaaactttatcgttgatattctag	1680	
DB	1621	AGCCTGTCAAACCAATGGTCTCCCAACGGCAGATGAACTTTATTCGTTGATATCTTAG	1680	
QY	1681	ataacgatttaacgtatgcagacattacatccttggagccattttaaaacaaatctctca	1740	
DB	1681	ATAACGATTTAACGTTATGCAGACATTACATCTCTTTGAGCCATTATTAAACAATCTCTCA	1740	
QY	1741	aggatccttgatgcagagagagctgttaacagttaccattcaagaagacacctgtacaagtac	1800	
DB	1741	AGGATCTTGATGCAGAGAGGCTGTAACTACCATCAAGGAAGACACTGTACAAGTAC	1800	
QY	1801	csagtggcagtagggcccgccgaagaagtgcgaacggaagaattgatgcaactcccaagag	1860	
DB	1801	CAGTGGCAGTAGGGCCCGCCGAAGAAGTGCCAAACGGAAGATTGATGCAACTCCAAGAGG	1860	
QY	1861	acqatttcgaattagaagaactgcagaagctccagaggaaggaagaattagattagaag	1920	
DB	1861	ACGATTTCCAAATTAGAAGGAACCTGCAGAAGCTCCAGAGGAAGGAATTAGTATTAGAG	1920	
QY	1921	gagaaggaagaaccaaaggaagaagccaagaagagagccaagaagaaggaagtgc	1980	
DB	1921	GAGAAGGAGAACCACCGAAGAGAGGCCAAGAGAGAGGAGCCACAGAGGAGAGGTGC	1980	
QY	1981	cagaagaagaatttagaggcaactccagaggacgatttcgaattagaagaaccaaagag	2040	
DB	1981	CAGAAGAAGAAATTAGAGGCAACTCCAGAGGACGATTTTCAATTAGAAGAAACCAACAGAGG	2040	
QY	2041	aagaagtagaagaacccgtagaggcggaagaactgcagaaggaagaagtggaagag	2100	
DB	2041	AAGAAGTAGAAGAAACCCGTAGAGGGCGAAGAAACTGCAGAAGGAGAAAGTGGAGAGG	2100	
QY	2101	tacctgcagaagtagaagaagtgggaagaggtacctgcagaagtagaagaagtgggaagag	2160	
DB	2101	TACCTGCAGAGTAGAAGAGTGAAGAGGTACCTGCAGAACTAGAGAGTGAAGAGG	2160	

QY 361 cccccgaagatgactttgcttaacagatttgccaattgaaacgcatgaggagtcacacg 420
|||||
Db 361 CCCCCGAAGATGACTTTGCTCTTAACAGATTGTCCTTAAGACGATGAGGAAGTCAACG 420
QY 421 aaacgtagatgagtgatcatattagagagaggtttccactgaagatatggaaacagaag 480
|||||
Db 421 AAACGTTAGATGGAGTGAATCATATTAGGAGAGGTTTCCACTCAAGATATGGAACAGAA 480
QY 481 atggtctcaacagatgatacggaacagaagaagactacctggtgatatggaggagaag 540
|||||
Db 481 ATGGCTCAACAGATGATACGGAACACAGAGAGGACTACTGCTGATATGGAAGGAGAG 540
QY 541 aagaagctggcgatattggaagcagggggaagaagctggtgatttgggaagcagggggaagaa 600
|||||
Db 541 AAGAAGCTGGCGATATGGGAAGCAGGGGAAGAGCTGGTGATTGGAAGCAGGGGAAGAAA 600
QY 601 ctggcatttgaagcaggggaagaacaggcagatttgaagcagggggaagagctgggtg 660
|||||
Db 601 CTGGCGATTGGAAGCAGGGGAAGAACTGGCGATTGGAAGCAGGGGAAGAGCTGGGTG 660
QY 661 atttgggaagcaggggaagaactggcgatttgggaagcaggggaagaactggagatggcg 720
|||||
Db 661 ATTTGGAAGCAGGGGAAGAACTGGCGATTGGAAGCAGGGGAAGAACTGGGAGTCCG 720
QY 721 aaactgaagaagagcaactggagatgcggaactgaactgaactggagcaactggtatgtatg 780
|||||
Db 721 AAACCTGAAGAAGAGCAACTGGAGATGCGGAACCTGAAATGCAGCAACTGCTATGTAG 780
QY 781 acacagaatagttcagctgatggagcagaaagacacatattctctctcaagaataatg 840
|||||
Db 781 ACACAGAAGATAGTTTCAAGTATGGAGCAGAAAGATACATGTTCTGCTCAAGAAATATG 840
QY 841 tacaactgcggatagtaatgatgcctctttggagagttttggataaagataataatt 900
|||||
Db 841 TACAACCTGCGGATAGTATGATGCTCTTTTGGAGATATTGGAATTAAGATATAATTT 900
QY 901 ttgatcatattaaagatttcgaagccactattcgaaacaaatttgcggtactgctaaac 960
|||||
Db 901 TTGATCATATTAAAGATTTTCGAGCCACTATTGCAACAAATTTGCGGGTACTGCTTAAC 960
QY 961 atgttacgggacaagaatttgcgaatgaacactgtaccattaccagtggcagagaagcccg 1020
|||||
Db 961 ATGTTACGGGACAAGAAATGCCCAATCAACCTGTACCATTAACGATGGCAGAGACCCG 1020
QY 1021 cgcaatccagcgggaagaattagatgcctccagagagatgacttcgcattagatggtta 1080
|||||
Db 1021 CGCAAGTACCAGCGGGAAGAAATTAGATGCCACTCCAGAGGATGACTTCGCATTTAGATGTTA 1080
QY 1081 cagaactcccgaggagtagaatttagtattagatgaagagcaactgaagaagaatcaa 1140
|||||
Db 1081 CAGAAATCTCCGAGGAAGTAGAAATTAGTATTAGATGAAGAGGCAACTGAAGAAGAAATCAA 1140
QY 1141 cggaagtgggaccacaggaagagccaccgaagaatttagatgcactccagagagtg 1200
|||||
Db 1141 CGGAAGTGGGACCACCGGAGAGAGGACCAACCCGAAATTAGATGCTCCAGAGGATG 1200
QY 1201 gatttcgcatagcgaactcagagaagagagaacacgaagaacgtagagggagaagaagaa 1260
|||||
Db 1201 GATTTCCGATTACACGAAACTGCAGAAGGAGAAACAGAAACGTFAGAGGGGAAGAAA 1260
QY 1261 cagaagaagctcagaagaagtagtatcagaagaactccagaagagagaagaagcttag 1320
|||||
Db 1261 CAGAAGAAGCTGCAGAAGGAGAAAGTATCAGAAGAACTCCAGAAGGAGGAAGAGGTTAG 1320
QY 1321 aggcaactcagagagatgatttcgcattagatggaactacattagaagaaaccgaagaaa 1380
|||||
Db 1321 AGGCAACTCCAGAGGATGATTTCCCAATTAGTGGAACTACATTTAGAAAGAAACCGAAGAAA 1380
QY 1381 ctgcagaagagagaagaaacgtagaggggaagaacccgtagagggagaagaacccgtag 1440
|||||
Db 1381 CTCGAGAAGGAGAAACCGTAGAGGGAGAGAAACCCGTAGAGGGGAGAAACCCGTAG 1440

QY 1441 agggagaagaagctgcagaagagagaagagttagaggaactccagaggatgacttcc 1500
|||||
Db 1441 AGGGAGAAGAAGCTGCAGAAGGAGAAGAGAGTTAGAGCAACTCCAGAGGATGACTTCC 1500
QY 1501 aattagaagaaccatcagagaagagaaggggaagagaaggaaggaaggaaggaag 1560
|||||
Db 1501 AATTAGAAGAACCATCAGGAGAAGGAGAGGGGAAGGAGAGGGAAGGGAAGGAGAAG 1560
QY 1561 gagaagcgttagtagcagtgccagtagtggccgaacccggtagaagtagtactcctgctc 1620
|||||
Db 1561 GAGAAGCGTTAGTAGCAGTGCAGTAGTGGCCGAACCCGGTAGAAGTAGTACTCTGCTC 1620
QY 1621 agcctgtcaaaccaatgctgcgtcccaacggcagatgaaactttatctgtgatatcttag 1680
|||||
Db 1621 AGCCTGTCAAAACCAATGCTGCTCCCAACGGCAGATGAAACTTTATTTCGTTGATATCTTAG 1680
QY 1681 ataacgatttaacgtatcagacattacatcctcttgagccattatttaacaaactccca 1740
|||||
Db 1681 ATAACGATTTAACGATATGCAGACATTTACATCTCTTTGAGCCATTTTAAACAAATCTCTCA 1740
QY 1741 aggtacctgatgcagagagggctgtaacagtgaccatcaaaaggaagcacctgtcacagtac 1800
|||||
Db 1741 AGGATCCTGATGCAGAGAGGCTGTAACAGTACCATCAAAAGGAAGCACCTGTACAAGTAC 1800
QY 1801 cagtggcagtagggccgcgcaagaagtgcacaacggagaagaattgatgcaactcccaagagg 1860
|||||
Db 1801 CAGTGGCAGTAGGGCCCGCGCAAGAAGTGCCAAACGGAAGAATTGTATCAACTCCAAGAGG 1860
QY 1861 acgatttcgaattagaagaactgcagaagctccagaggaaggaatagattagattagaag 1920
|||||
Db 1861 ACGATTTGCAATTTAGAGGAACCTGCAGAACTCCAGAACTCCAGAGGAAGAGANTAGTATTAGAAG 1920
QY 1921 gagaagggagaacccaggaagagggccaagagaagggagccacaagaagaggaagtgc 1980
|||||
Db 1921 GAGAAGGAGAACCAACGGAAGAAGAGCCAAAGAAAGAGGAGAGCAACAGAAAGGAGAAGTGC 1980
QY 1981 cagaagaagaattagaggcaactccagaggagcatttcgaattagaagaaccacaaggaag 2040
|||||
Db 1981 CAGAAGAAGAATTTAGAGGCCAACTCCAGAGGACGATTTGCAATTTAGAAGAACCACAGAGGAG 2040
QY 2041 aagaagtgaagaagaacccgtagaggcggaagaactgcagaaggaagaagtggaagagg 2100
|||||
Db 2041 AAGAAGTAGAAGAAACCGTTAGAGGGCGAAGAACTGCAGAAAGGAGAAGAGTGAAGAGG 2100
QY 2101 tacctgcgaagaatcagaagaagtgggaaggggtacctgcgaagaatagaagaagtgggaagagg 2160
|||||
Db 2101 TACCTGCGAAGTAGAAGAAGTGGAAAGGTGAGGAGGTACCTGCAAGTAGAAGAAGTGAAGAGG 2160
QY 2161 taccagaagaagttagaagaaggtaccgcgagaagttagaagaagtgggaaggggtaccagaag 2220
|||||
Db 2161 TACCAAGAAGTAGAAGAGGTACCCGCAAGAGTGAAGAGTGAAGAGTGAAGAGGTACCAAGAG 2220
QY 2221 asgtggaagagggtaccagaagaagtgggaaggggtaccagaagaagtgggaagagggtaccag 2280
|||||
Db 2221 AAGTGGAAAGAGGTACCAGAAGAAGTGGAAAGAGGTACCAGAAGAAGTGGAAAGAGGTACCAAG 2280
QY 2281 aagaagtggaagaagtggaagaagttagaagaagttagaggtaccagcggtagtagaagtag 2340
|||||
Db 2281 AAGAAGTGGAAAGTAGAAGAAGTGGAAAGAGTAGAAGAGTAGTACCGGTAGTAGAAGTAG 2340
QY 2341 agttaccagcggtagtagaagaagaggtgcccagaagaagttagaagaagaagaagaagagg 2400
|||||
Db 2341 AAGTACCAGCGGTAGTAGAAGAAGAGGTCGCCAGAAGAGTAGAAGAAGAAAGAAAGAGGAGG 2400
QY 2401 aagaaccagtagaggaagaagatgtattacaatttagtaataccatcggaagaagataac 2460
|||||
Db 2401 AAGAACCCAGTAGAGGAGAGAGATGTTATACATTTAGTATAGTATCCATCGGAAGAGATATAC 2460
QY 2461 aattagacaaacccaagaagaagacgaatttaggctctggaattttatctatcatcgacatgc 2520
|||||
Db 2461 AATTAGACAAACCAAAAGAAAGACGAATTAGGCTCTGGAATTTTATCTATCATCGCATGC 2520
QY 2521 actaccaagacgttccaaaagggaatttatggaagaagaagaagaagaagtagtaccat 2580

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 96243720
FILING DATE: 13-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KILYK JR., JOHN
REGISTRATION NUMBER: 30763
REFERENCE/DOCKET NUMBER: 72882
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE: 5'-end is phosphorylated
US-09-065-058-16

Query Match 0.8%; Score 28; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 0.021; 0; Indels 0; Gaps 0;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 200 ttttttttttttttttttttttttttcta 227
Db 49 TTTTtttttttttttttttttttttttttcta 76

RESULT 6
US-08-828-007-7
Sequence 7, Application US/08828007
Patent No. 5866412
GENERAL INFORMATION:
APPLICANT: Chen, Hong
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC
DISORDERS
TITLE OF INVENTION: DISORDERS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828,007
FILING DATE: 27-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-069
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Antisense oligonucleotide
LOCATION: 1...50
OTHER INFORMATION:
US-08-828-007-7

Query Match 0.8%; Score 27; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.054; 0; Indels 0; Gaps 0;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 195 aaaaattttttttttttttttttttttttt 221
Db 24 AAAAattttttttttttttttttttttttttt 50

RESULT 7
US-07-867-106-4
Sequence 4, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved plasmid vectors for cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3138 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
US-07-867-106-4

Query Match 0.8%; Score 27; DB 1; Length 3138;
Best Local Similarity 100.0%; Pred. No. 0.044; 0; Indels 0; Gaps 0;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 195 aaaaattttttttttttttttttttttttt 221

Db 1938 AAAAAATTTTTTTTTTTTTTTTTTTTTTTT 1964
|||||

RESULT 8
US-09-244-794A-1/c
; Sequence 1, Application US/09244794A
; Patent No. 6214553
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350006
; CURRENT APPLICATION NUMBER: US/09/244,794A
; CURRENT FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: 60/035,963
; PRIOR FILING DATE: 1997-01-27
; PRIOR APPLICATION NUMBER: 60/064,491
; PRIOR FILING DATE: 1997-11-06
; PRIOR APPLICATION NUMBER: 09/007,005
; PRIOR FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 76
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
US-09-244-794A-1

Query Match 0.8%; Score 26; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttc 225
|||||

Db 69 TTTT

RESULT 9
US-09-247-190-1/c
; Sequence 1, Application US/09247190
; Patent No. 6261804
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350005
; CURRENT APPLICATION NUMBER: US/09/247,190
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-21
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 76
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
US-09-247-190-1

Query Match 0.8%; Score 26; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttc 225
|||||

Db 69 TTTT

RESULT 10
US-08-150-331-45/c
; Sequence 45, Application US/08150331
; Patent No. 5516512
; GENERAL INFORMATION:
; APPLICANT: DORSSERS J., LAMBERTUS C.
; APPLICANT: VAN LEEN, ROBERT W.
; TITLE OF INVENTION: MUTANTS OF HUMAN INTERLEUKIN-3
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/150,331
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/651,437
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: GRACEY, NANCY J.
; REGISTRATION NUMBER: 28,216
; REFERENCE/DOCKET NUMBER: 24615-20010.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 596 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..435
US-08-150-331-45

Query Match 0.8%; Score 26; DB 1; Length 596;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttc 225
|||||

Db 561 TTTT

RESULT 11
US-08-467-559B-1
; Sequence 1, Application US/08467559B
; Patent No. 5928890
; GENERAL INFORMATION:
; APPLICANT: LI, YI
; TITLE OF INVENTION: HUMAN AMINE RECEPTOR

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: UNITED STATES OF AMERICA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,559B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.084.0000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 252..1262
US-08-467-559B-1

Query Match 0.8%; Score 26; DB 2; Length 1380;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 aaaaattttttttttttttttttttttt 221
|||||
DB 158 AAAATTTT

RESULT 12
US-08-396-479B-7/C
Sequence 7, Application US/08396479B
Patent No. 5612455
GENERAL INFORMATION:
APPLICANT: HOEY, Timothy
TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/396,479B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-59450-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 210 277299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2406 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 211..2337
US-08-396-479B-7

Query Match 0.8%; Score 26; DB 1; Length 2406;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 tttttttttttttttttttttttttttttt 225
|||||
DB 2398 TTTT

RESULT 13
US-08-818-823-7/C
Sequence 7, Application US/08818823
Patent No. 5708158
GENERAL INFORMATION:
APPLICANT: HOEY, Timothy
TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,823
FILING DATE: 14-MAR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/396,479
FILING DATE: 02-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59450-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 210 277299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2406 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 211..2337
US-08-818-823-7

Query Match 0.8%; Score 26; DB 1; Length 2406;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 tttttttttttttttttttttttttttttttg 225
|||||
Db 2398 TTTT

RESULT 14
US-09-740-035-3
; Sequence 3, Application US/09740035
; Patent No. 6344353
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001058
; CURRENT APPLICATION NUMBER: US/09/740,035
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 19736
; TYPE: DNA
; ORGANISM: Human
US-09-740-035-3

Query Match 0.8%; Score 26; DB 4; Length 19736;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 aaattttttttttttttttttttttttttttttg 222
|||||
Db 3500 aaatttttttttttttttttttttttttttttg 3525

RESULT 15
US-09-328-111-586
; Sequence 586, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 586
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-328-111-586

Query Match 0.7%; Score 25; DB 4; Length 337;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 tttttttttttttttttttttttttttttgtt 224
|||||
Db 10 tttttttttttttttttttttttttttttgtt 34

RESULT 16
US-09-385-982-342
; Sequence 342, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 342
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (1)...(342)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-342

Query Match 0.7%; Score 25; DB 4; Length 342;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 tttttttttttttttttttttttttttttgtt 224
|||||
Db 9 tttttttttttttttttttttttttttttgtt 33

RESULT 17
US-08-628-413-1/C
; Sequence 1, Application US/08628413
; Patent No. 5834192
; GENERAL INFORMATION:
; APPLICANT: Akerblom, Ingrid E.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN CACHEXIA ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,413
; FILING DATE: Filed Herewith

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0062 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: BRSTTUT01
; CLONE: 607227
; US-08-628-413-1

Query Match 0.7%; Score 25; DB 2; Length 523;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 aaattttttttttttttttttttttttt 221
| | | | | | | | | | | | | | | | | |
DB 477 AAATTTTTTTTTTTTTTTTTTTTTT 453

RESULT 18
US-08-131-365B-41/c
; Sequence 41, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 547 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
```

```
; DESCRIPTION: /desc = "DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..111
; US-08-131-365B-41

Query Match 0.7%; Score 25; DB 1; Length 547;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 tttttttttttttttttttttttttt 224
| | | | | | | | | | | | | | | | | |
DB 540 TTTTTTTTTTTTTTTTTTTTTT 516

RESULT 19
US-08-668-123-41/c
; Sequence 41, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 547 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..111
; US-08-668-123-41

Query Match 0.7%; Score 25; DB 2; Length 547;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 200 tttttttttttttttttttttttttttttttgtt 224
|||||
Db 540 TTTT

RESULT 20
US-09-385-982-427
; Sequence 427, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1999-06-08
; EARLIER FILING DATE: 1999-01-27
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 427
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(567)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-427

Query Match 0.7%; Score 25; DB 4; Length 567;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 tttttttttttttttttttttttttttttttgtt 224
|||||
Db 14 tttttttttttttttttttttttttttttttgtt 38

RESULT 21
US-08-313-681A-3/C
; Sequence 3, Application US/08313681A
; Patent No. 5618675
; GENERAL INFORMATION:
; APPLICANT: Larrick, James W.
; APPLICANT: Wright, Susan C.
; APPLICANT: Hirata, Mishimasa
; TITLE OF INVENTION: Human Cationic Proteins Having
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/313,681A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.

; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15325-9-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 587 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-313-681A-3

Query Match 0.7%; Score 25; DB 1; Length 587;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 tttttttttttttttttttttttttttttttgtt 224
|||||
Db 585 TTTT

RESULT 22
US-09-322-911-3/C
; Sequence 3, Application US/09322911
; Patent No. 6103888
; GENERAL INFORMATION:
; APPLICANT: Larrick, James W.
; APPLICANT: Wright, Susan C.
; APPLICANT: Hirata, Mishimasa
; APPLICANT: Balint, Robert F.
; TITLE OF INVENTION: Human Cationic Proteins Having
; TITLE OF INVENTION: Lipopolysaccharide Binding and Anti-Coagulant Activity
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/322,911
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/691,280
; FILING DATE: August 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/916,761
; FILING DATE: July 17, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/916,765
; FILING DATE: July 17, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06731
; FILING DATE: July 15, 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/313,681
; FILING DATE: September 27, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 15325-000920
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400

TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 587 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-322-911-3

Query Match 0.7%; Score 25; DB 3; Length 587;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttt 224
|||||
Db 585 TTTTtttttttttttttttttttttttttttttt 561

RESULT 23
US-09-385-982-262
; Sequence 262, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 262
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(593)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-262

Query Match 0.7%; Score 25; DB 4; Length 593;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttt 224
|||||
Db 6 ttttttttttttttttttttttttttttttt 30

RESULT 24
US-09-385-982-288
; Sequence 288, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393

EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 288
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(607)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-288

Query Match 0.7%; Score 25; DB 4; Length 607;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttt 224
|||||
Db 6 ttttttttttttttttttttttttttttttt 30

RESULT 25
US-09-385-982-389
; Sequence 389, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 389
; LENGTH: 629
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(629)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-389

Query Match 0.7%; Score 25; DB 4; Length 629;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttt 224
|||||
Db 5 ttttttttttttttttttttttttttttttt 29

RESULT 26
US-09-352-990-3/c
; Sequence 3, Application US/09352990
; Patent No. 6255090
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase

Query Match 0.7%; Score 25; DB 3; Length 770;

```

US-08-441-629-1/C
; Sequence 1, Application US/08441629
; Patent No. 5766923
;
; GENERAL INFORMATION:
;
; APPLICANT: Kirschner, Marc W.
; APPLICANT: Kinoshita, No. 5766923/yuki
; TITLE OF INVENTION: RECEPTOR-LIGAND ASSAY
;
; NUMBER OF SEQUENCES: 17
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,629
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;

```

Search completed: June 14, 2002, 20:29:19
Job time: 12413 sec